DOCUMENT RESUME

ED 148 834 TM 006 516

TITLE	On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline and Appendices.
INSTITUTION	College Entrance Examination Board, New York, N.Y.
SPONS AGENCY	Educational Testing Service, Princeton, N.J.
PUB DATE	77
NOTE	837p.; For individual appendices, see TM 006 750-776; For related document, see TM 006 805
AVAILABLE FROM	College Board Publication Orders, Box 2815, Princeton, New Jersey 08541 (Report, Item No. 251637, $4.00, appendices, Item No. 2517027, $25.00, also available separately)
EDRS PRICE	MF-$1.50 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; College Bound Students; *College Entrance Examinations; Educational Change; Educational Practice; *Educational Problems; *Educational Trends; Family Influence; *Performance Factors; Population Trends; *Scores; Secondary Education; Social Factors; Student Characteristics; Student Motivation; Testing Problems
IDENTIFIERS	*Scholastic Aptitude Test; *Test Score Decline

ABSTRACT

After two years of investigation, a panel appointed to advise the College Entrance Examination Board and Educational Testing Service on the recent change in Scholastic Aptitude Test (SAT) scores concluded that the 14-year score decline is a complex phenomenon, yielding neither simple explanations nor easy solutions. As a result of looking at the technical aspects of the SAT score decline, the panel determined that the drop does not result from the test having become harder. A larger part of the decline between 1963 and 1970 was traced to compositional changes in the college-bound population. As compared to 15 or 20 years ago, a broader cross section of American youth are taking the test. The panel identified six other developments that have also contributed to this decline: proliferation of elective courses, particularly in the area of English and verbal skills; a diminished seriousness of purpose in the learning process as reflected in tolerance of excessive absenteeism, grade inflation, easier textbooks, and generally lowered standards; increased viewing of television; changing structure of the family; social turbulence of the 1960's and 1970's; and decreased learning motivation. (Author/EVH)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED148834

On Further Examination

Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline

College Board

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

MICROFICHE ONLY

College Entrance
Examination Board

TM006 516

On Further Examination

Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline

---

Willard Wirtz, *chairman*
Harold Howe II, *vice chairman*
Bernard C. Watson
Ralph W. Tyler
Ledyard R Tucker
Vivian H. T. Tom
Robert L. Thorndike
Barbara Thompson
Thomas W. F. Stroud
Rosedith Sitgreaves
Wilbur Schramm
Katherine P. Layton
Owen B. Kiernan
H. Thomas James
Matina S. Horner
Edythe J. Gaines
Frank W. Erwin
Bruce K. Eckland
Luis C. Cortes
Sandra A. Clark
Benjamin S. Bloom

College Entrance Examination Board
New York, 1977

Copies of this book may be ordered from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. The price is $4. For information about purchasing the Appendixes, which are available separately as well as in one volume, write to College Board Publication Orders, Box 2815, Princeton, New Jersey 08540.

Editorial inquiries concerning this book should be directed to Editorial Office, College Entrance Examination Board, 888 Seventh Avenue, New York, New York 10019.

Copyright © 1977 by the College Entrance Examination Board. All rights reserved.
Library of Congress Catalog Card Number: 77-84842
Printed in the United States of America

Prefatory Note

"No topic related to the programs of the College Board has received more public attention in recent years than the unexplained decline in scores earned by students on the Scholastic Aptitude Test. The trustees and the officers of the College Board believe that we must do all that we can to investigate and interpret this phenomenon to the public at large. . . .

"We are appointing a blue-ribbon panel to assist in making sense out of the complex and interrelated issues involved. The panel will be asked to audit the steps already taken to insure the psychometric integrity of the tests, to suggest additional ones if appropriate, to examine other kinds of research already done, and to identify research that still needs to be done in order to deal effectively with the score decline issue as it relates to candidate population, secondary education, and society."

With these words, written in October 1975 in my capacity as president of the College Board, and in consultation with William W. Turnbull, president of Educational Testing Service, I invited the 21 members of the Advisory Panel on the Scholastic Aptitude Test Score Decline to undertake this work. They were asked to function "as an advisory body to the presidents of the College Board and Educational Testing Service," and in that capacity "to consider the matter of the SAT score decline and assist in the [development of an] understanding of it."

The panelists have been extraordinarily generous in the investment of their personal time and effort, and they have brought to the problem a level of professional expertise that would be hard to match. In addition to the four lengthy plenary sessions held by the full panel and the several meetings of its three working subgroups, uncounted hours of homework and individual exchanges of correspondence surrounded the deliberative process. The discussions at each session were based on diligent advance attention to a wealth of background materials requested by the members and produced by internal and external researchers. The panel's deliberations were also enriched by the availability of a number of contemporary external documents dealing with various aspects of the score decline issue. About two-thirds of the costs of the panel-generated investigations, as well as the expenses of the panel itself, have been borne by the College Board, with the rest underwritten by Educational Testing Service.

Although the panel was technically appointed to advise the College Board and Educational Testing Service, its mission from the start emphasized its independence, including the freedom to deliberate such issues as it chose to consider and to report its findings in the public interest, whether or not they might be critical of the College Board or ETS. Consistent with that spirit, the College Board is pleased to publish the panel's report as it was submitted to us. In doing so, the College Board and ETS re-serve the right—indeed, recognize their responsibility—to comment over time on its findings as particularly interested members of the educational community and the greater society we seek to serve. Meanwhile, I wish to express President Turnbull's and my appreciation to the members of the panel and to commend their report to the attention of all who care about education in the United States.

S. P. Marland, Jr.
July 1977

Contents

Acknowledgments. vi
Introduction 1
Part One. The Scholastic Aptitude Test and the Score Decline 3
Part Two. An Unchanging Standard. 8
  As It Was in the Beginning 8
  A Changing Context? 10
Part Three. The Two Score Declines. 13
  Compositional Change. 13
  Pervasive Change 18
  Through Other Looking Glasses 21
Part Four. Circumstantial Evidence 25
  In the Schools 25
    Courses of Study 25
    Learning Standards 28
    Staff and Facilities. 31
  The Broader Learning Context. 33
    Parents as Teachers 33
    Television – Test of the Modern World. 35
    A Decade of Distraction 37
  Motivation 38
  In the View of Others 40
Part Five. Summing Up 44
Reference Notes 49
Annotated List of Studies and Papers 55
Bibliography 64
Sample Questions from the SAT 71
Advisory Panel on the Scholastic Aptitude Test Score Decline. 75

Acknowledgments

The panel's debts are so many that simple acknowledgment of the largest of them will have to suffice.

Edmund J. Farrell and Alfred L. Putnam, chairmen of the College Board's Discipline Committees in English and Mathematics, respectively, served in effect as members of the panel, taking full part in our discussions and providing special insights and assistance.

George H. Hanford, senior vice president of the College Board, contributed immeasurably to the panel's functioning. The delicate role he played as chief of staff and intermediary was vital to our effort to be both fully informed and objectively critical. Yvonne Wharton worked closely with him in all of this, as did Lillian Tucci.

Paul Barton, senior associate at the National Manpower Institute, shared the responsibilities of the panel's chairman. Arlene Huff helped draft after draft through its dreariest stage.

The officers and staffs of the College Board and ETS provided us with assistance in extraordinary measure.

We asked a number of people outside the panel for information and for reactions and are indebted to them for their response. We have also benefited greatly from the communications received from several hundred individuals who have written simply in expression of their concerns about this subject.

Willard Wirtz, *chairman*

On Further Examination

Introduction

Every year, for 14 years now, there has been a drop in the average scores more than a million high school juniors and seniors get on the Scholastic Aptitude Test (SAT) they take in seeking admission to college. The panel has been asked by the College Board (which sponsors the SAT) and Educational Testing Service (ETS) (which develops and administers it) to look into this situation and to report publicly its findings and conclusions about the score decline. This is that report.

Starting with the technical aspects of the score decline as they appear under microscopic scrutiny and seem to permit objective analysis, we try in the later parts of the report to provide broader perspective. This means proceeding from what we have been able to establish on the basis of available data, to what we believe is reasonable interpretation of broader evidence, to what we speculate about as concerned individuals.

The particular interest of the College Board and ETS is in determining whether anything about the test itself has contributed to the decline in scores. Following a factual statement in Part One, we deal with this set of questions in Part Two.

The public's interest however, is not in the psychometric technicalities of the SAT score decline but in its implications regarding what is widely perceived as serious deterioration of the learning process in America. More and more high school graduates show up in college classrooms, employers' personnel offices, or at other common checkpoints with barely a speaking acquaintance with the English language and no writing facility at all. Parents watch children come home from school, without homework, to sit passively hour after hour and day after day in front of television sets until they have spent more time there than anyplace else except in bed. Although the SAT score figures are too small a window for surveying this broad condition, they provide special insight into it.

A major part of the decline in college entrance examination scores, especially in its earlier stages, is clearly traceable to a change in the composition of the student group taking the test, resulting from the deliberate and historic decision in this country in the 1960s to extend and expand educational opportunity and to eliminate previous discrimination in according it. The SAT data permit a relatively objective determination—covered in Part Three of the report—of the effects on the test scores of the changing composition of the group taking these examinations.

Equally clearly, the score decline also reflects, particularly in the 1970s, the operation of other more pervasive forces. Although the test score statistics do not themselves indicate the nature of these forces, there are available data that permit informed conjecture about the impact on learning of changes that have taken place in both the schools and the society during a period of turbulence and distraction rarely paralleled in American history. This context is described and discussed in Part Four.

The sources on which we have relied are indicated in Reference Notes at the end of the report. Many of these sources are special studies made for the panel, and they are available in a separate volume of Appendixes to the report.

No one of us on the panel would have put everything in the report in the form it takes here. We share the feeling, nevertheless, that this comes close enough to what seems important to all of us to warrant submerging our small differences so that our larger agreement is plain.

Part One. The Scholastic Aptitude Test and the Score Decline

THE Scholastic Aptitude Test has been used since the 1920s to help determine high school students' apparent preparedness for college. Many colleges and universities require applicants for admission to submit SAT scores for consideration along with high school academic and extracurricular records, letters of recommendation, and the results of personal interviews. Others make similar use of a comparable examination administered by the American College Testing (ACT) Program. Still others require no standardized college entrance examination.

Given in different editions several times a year at locations all over the country, the SAT is taken primarily by seniors but also by a large number of juniors and by a few others. Some students take it more than once. The roughly one million young people in each high school senior class who take the SAT represent approximately a quarter of their age group as a whole and about half of the number going on to college. The figures for the past 26 years are in Table 1.

The SAT includes a Verbal and a Mathematical part. Scores for both parts are computed and reported separately on a scale of 200 to 800. Two and a half hours are allowed for taking the test. (The time, originally three hours, was shortened in 1973 to permit adding the 30-minute Test of Standard Written English; but the results on the latter test are not included in the reported SAT scores.)

The Mathematical portion of the SAT, which requires as background mathematics typically taught in grades one through nine, depends less on formal knowledge than on reasoning; it measures students' problem-solving ability in three areas—arithmetic reasoning, elementary algebra, and geometry. The Verbal portion, designed to assess reading skills and understanding of word relationships, covers four areas—antonyms, analogies, sentence completion, and reading comprehension; the material for this test is drawn from social, political, scientific, artistic, philosophical, and literary writing. Sample questions from the SAT are included at the end of the report.

The panel has reviewed the SAT score pattern as it has developed over the past 26 years. These figures are in Table 2 and Chart A. For convenience, we refer throughout the report to the year in which an academic or SAT year *ends*—to 1952, as an example, for 1951-52. Wherever we refer to the number of SAT takers or the average scores in a particular year, the reference will be, for all years since 1966, to the *number or the scores of high school seniors taking the test*—as distinguished from the *total number of tests taken that year* (which was the basis on which the records were kept prior to 1967); this permits better comparisons between particular groups, or "cohorts" of students. Finally, when we refer to SAT "scores" the reference is to what are technically "scaled scores."

Table 1. Numbers of 18-year-olds, high school graduates, first-time, degree-credit enrollments, and SAT takers, 1951-52 to 1976-77 (in thousands)

	18-year-olds[1]	High school graduates[2]	First-time, degree-credit enrollment[3]	SAT tests taken[4]
1951-52	2,058.0	1,196.5	532.3	81.2
1952-53	2,160.0	1,198.3	566.0	95.5
1953-54	2,135.0	1,276.1	624.9	118.1
1954-55	2,142.0	1,351.0	668.1	154.5
1955-56	2,244.0	1,414.8	715.0	208.6
1956-57	2,274.0	1,439.0	721.5	270.5
1957-58	2,307.0	1,505.9	772.3	376.8
1958-59	2,431.0	1,639.0	818.3	469.7
1959-60	2,613.0	1,864.0	923.1	564.2
1960-61	2,976.0	1,971.0	1,018.4	716.3
1961-62	2,816.0	1,925.0	1,030.6	802.5
1962-63	2,786.0	1,950.0	1,046.4	933.1
1963-64	2,763.0	2,290.0	1,224.8	1,163.9
1964-65	3,804.0[5]	2,665.0	1,441.8	1,361.2
1965-66	3,536.0	2,632.0	1,378.0	1,381.4
1966-67	3,545.0	2,679.0	1,439.0	1,422.5
1967-68	3,539.0	2,702.0	1,629.8	1,543.8
1968-69	3,676.0	2,829.0	1,748.7	1,585.6
1969-70	3,780.0	2.896.0	1,780.1	1,605.9
1970-71	3,875.0	2,943.0	1,765.6	1,537.2
1971-72	3,970.0	3,006.0	1,740.4	1,459.9
1972-73	4,044.0	3,037.0	1,756.9	1,398.4
1973-74	4,093.0	3,069.0	1,854.4	1,354.0
1974-75	4,243.0	3,140.0	1,910.0	1,371.2
1975-76	4,253.0[6]	3,150.0(est.)	2,008.0(est.)	1,415.0
1976-77	4,239.0(est.)[6]	3.147.0(est.)	2,008.0(est.)	1,401.9

[1] Age as of July 1 of the later year. Estimates based on the April 1 decennial census surveys modified by records of births, deaths, immigration for intervening years. Estimates are for the total population, including Armed Forces overseas, the resident population, and the civilian population. (U.S. Bureau of the Census, *Current Population Reports*, Series P25, No. 311, 519, and 614, "Estimates of the Population of the United States by Age, Sex, and Race." U.S. Government Printing Office, Washington, D.C. 20402.)

[2] Includes regular public and nonpublic schools, residential schools for exceptional children, subcollegiate departments of institutions of higher education, federal schools for Indians, and federally operated schools on federal installations. Excludes equivalency certificates. Prior to 1960 data were collected only every other year, those ending in an even number. Beginning in 1959-60, Alaska and Hawaii are included. (National Center for Education Statistics, *Projections of Education Statistics*, annual. U.S. Government Printing Office, Washington, D.C. 20025.)

[3] Opening fall enrollment of first-time degree-credit students the following year. Beginning in 1960, Alaska and Hawaii are included. Beginning in 1953, extension students are included. (National Center for Education Statistics, *Digest of Education Statistics, 1975 edition*. U.S. Government Printing Office, Washington, D.C. 20402.)

[4] Figures listed are for number of SATS taken in a given testing year. Candidates may be from any grade and are counted more than once if they repeated the test (College Entrance Examination Board).

[5] This figure, which appears so high as to suggest error, is confirmed by reference to birth statistics (on comparable fiscal year basis) for 1945-46 and 1946-47.

[6] Information from National Center for Education Statistics, Department of Health, Education, and Welfare, June 28, 1977.

In 1952, the SAT-Verbal score average for all test takers was 476 and the SAT-Mathematical average 494. Although there were year-to-year fluctuations in these averages, they remained substantially level but moved up slightly through the 1950s and into the early 1960s; by 1963, the SAT-Verbal average was up 2 points to 478, the SAT-Mathematical average up 8 points to 502. Then in 1964, both score averages started dropping. They have dropped ever since, considerably more on the Verbal than on the Mathematical part of the test. The decline, relatively gradual through about 1970, became sharper after that, especially for the Verbal scores. The past two years have suggested a possible leveling out.

The panel has considered how far back to go in trying to analyze this scoring pattern. A 20-year comparison (1957 to 1977) would show about the same decline that a comparison of the 1963 and 1977 figures does. It is generally assumed that the increase in SAT scores in the early 1960s, especially the spurt that made 1963 a high year, may have reflected the results of the post-Sputnik acceleration of educational effort in this country; and this may have significance in appraising the present prospects, which we consider real, of a comparable recovery during the next few years.

The statistical evidence for that earlier period is exceedingly thin, however, except for the SAT scores themselves. We have accordingly concentrated on the 1963-to-1977 decline: the 49-point drop during this 14-year period in the score average on the Verbal part (from 478 in 1963 to 429 for 1977), and a 32-point drop (from 502 to 470) on the Mathematical part. The Mathematical decline is 31 points, if the comparison is made on an "All candidates" basis for both years.

How significant is this decline?

When the "standard deviations" involved here are taken into account, the decline in scores means that only about a third of the 1977 test takers do as well as half of those taking the SAT in 1963 did. But *how much* worse are students doing now than their counterparts used to do? Although this can't be answered with precision, a decline of this magnitude continuing over a 14-year period, following a previous period of stable or even slightly rising score averages, is clearly serious business.

The decline must be put, at the same time, in broader perspective.

Any generalization from the SAT statistics has to be carefully qualified. It should not be extended to cover the situation of American youth as a whole or the overall effectiveness of the learning process.

The college entrance examinations are represented by the agencies administering them only as indicating students' probable accomplishments in terms of college academic grades, particularly their first-year grades. Recently published College Board *Guidelines on the Uses of College Board Test Scores and Related Data* warn sharply against their misuse as measures of the broader effectiveness of elementary and secondary education in general.

The SAT figures cover only students who are still in high school at the 11th and 12th grades and who are considering going on to college, particularly to colleges or universities requiring applicants to take the SAT. This is a significantly different test-taking population, a different cross section of young people, from what it was 14 years ago. The score decline has taken place, furthermore, during a period of such

Table 2. Scholastic Aptitude Test Score Means, 1951-52 to 1976-77

Academic year	*SAT-Verbal*		*SAT-Mathematical*	
	All candidates[1]	*High school seniors*[2]	*All candidates*[1]	*High school seniors*[2]
1951-52	476		494	
1952-53	476		495	
1953-54	472		490	
1954-55	475		496	
1955-56	479		501	
1956-57	473		496	
1957-58	472		496	
1958-59	475		498	
1959-60	477		498	
1960-61	474		495	
1961-62	473		498	
1962-63	478		502	
1963-64	475		498	
1964-65	473		496	
1965-66	471		496	
1966-67	467	466	495	492
1967-68	466	466	494	492
1968-69	462	463	491	493
1969-70	460	460	488	488
1970-71	454	455	487	488
1971-72	450	453	482	484
1972-73	443	445	481	481
1973-74	440	444	478	480
1974-75	437	434	473	472
1975-76	429	431	470	472
1976-77	429	429	471	470

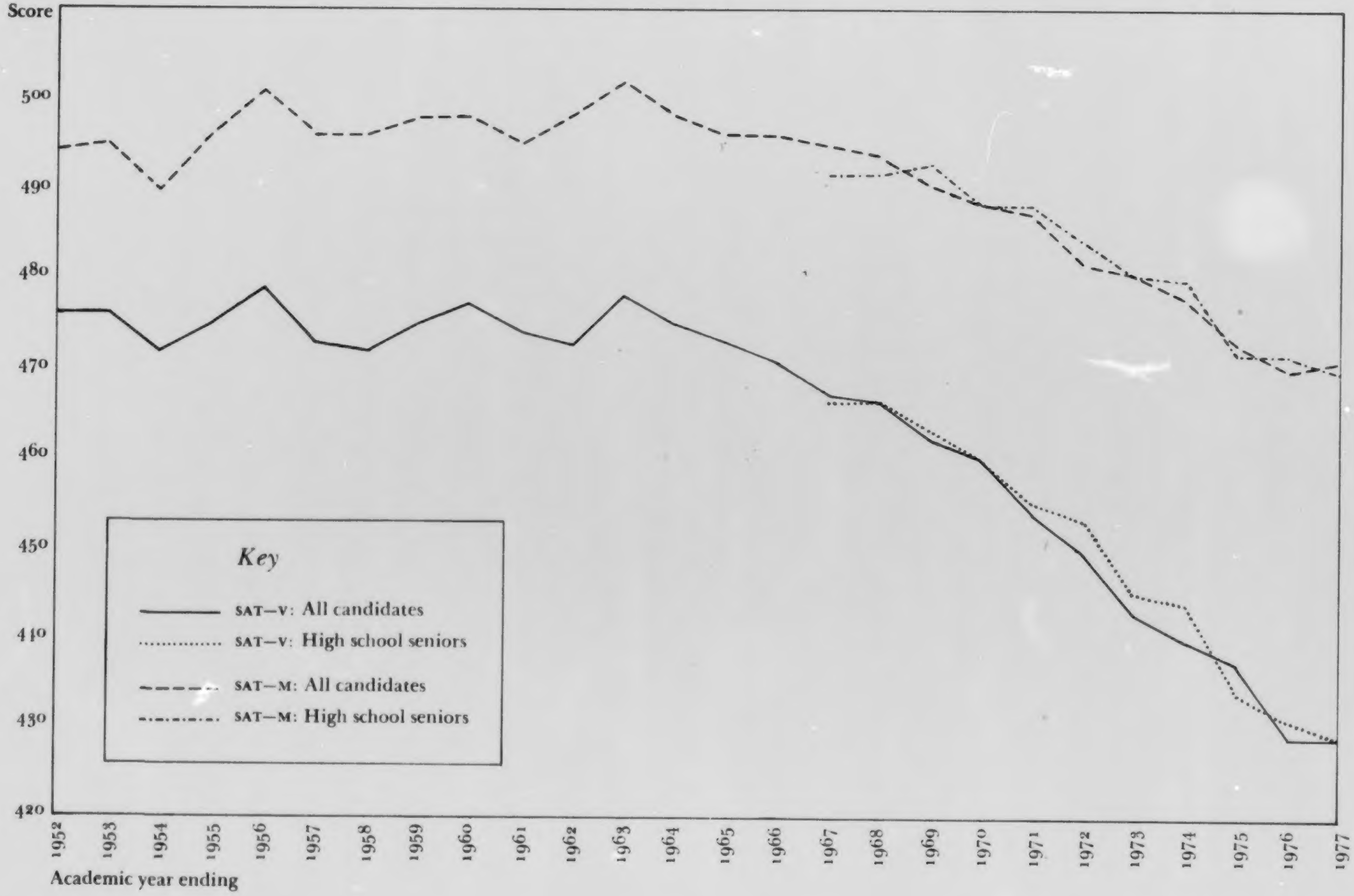
[1] An individual is counted as many times as he or she is tested.

[2] Each candidate is counted only once, using latest score earned (figures for 1967 through 1971 are estimates; for 1972 through 1977 are actual).

extraordinary national disruption, with particular effects on young people, that the change in the SAT score pattern is properly appraised only with full recognition of that context.

The report proceeds on the basis of the panel's persuasion that the score decline reflects a situation demanding serious attention but warranting, at the same time, cool-headedness about both the assessment of responsibility and the instruction this experience offers the future.

Chart A. *Scholastic Aptitude Test Scaled Score Means, 1952 to 1977*



Part Two. An Unchanging Standard

Few teachers worth the name ever settle two inner struggles: about the consistency of tests and grades with learning's ideals, and about whether to hold successive classes to a constant standard of measurement or to grade each of them "to the curve" – so that the proportion of high, medium, and low grades stays the same even though the overall quality level changes.

The SAT is designed to be an unchanging measurement. In technical terms, the score scales for the Verbal and Mathematical sections were established in 1941 and 1942, with mean scores of 500 and standard deviations of 100. Considerable effort has been made since that time to keep the test a sufficiently constant measure so that any particular score received on a current test indicates the same level of ability to do college work that the same score did 36 or 20 or 5 or 2 years ago. The SAT measures individual students' capacities not only in comparison with their peers in the particular group but also in comparison with those who took the test in earlier years.

This suggests two possible explanations of the SAT score decline involving the test itself. Because most of the SAT questions are changed with each new edition of the test, it could happen that the test would become inherently more difficult to score well on – "harder" than it used to be as a consequence of changes in the test questions and in the equating and scaling procedures involved in determining scaled scores.

An apparently converse possibility presents what is actually a very different type of issue. If the test were kept the same in every respect, but if those taking it were trained in different pursuits or by processes less consonant with the test, the resultant disjuncture would also present a possible explanation of a decline in the test scores. The point would be not that the test had changed but that by staying the same it had become less relevant to the training preceding it.

As It Was in the Beginning

The SAT score decline does not result from changes in the test or in the methods of scoring it.

Although we have gone to considerable lengths in exploring this possibility, there is no point in laboring the conclusion. The ETS procedures for "equating" successive editions of the test (by including in each new edition key questions from earlier tests and then "scaling" raw scores according to the performance on these questions) and for checking against "item obsolescence" are as sophisticated and reliable as the state of the psychometric art permits. We find complete agreement about this in the profession, and we have pressed the matter to the point of adequate lay understanding and concurrence.

These procedures neither pretend to nor do achieve perfection. In fact, two technical analyses of the equating and scaling practices, made at the panel's request (one comparing 1972 and 1960 data; the other, 1973 and 1963), indicate an "upward drift" of between 8 and 12 points in the scaling of the scores. In order to check this, a substantial group of high school students (3,174 from 66 schools) was given both 1963 and 1973 editions of the test, half of them taking the 1963 test first and half starting with the 1973 test. The results confirmed the earlier technical analysis: the test takers averaged higher scores on the 1973 tests than on the 1963 tests, both Verbal and Mathematical. This means that the declines in the ability the SAT measures have been from 8 to 12 points larger than the recorded and reported scores indicate. The panel considered casting its report in terms of figures adjusted to reflect this difference, but decided against this because of the impossibility of identifying the year or years in which the drift occurred.

A special inquiry was made into the possibility that the shortening of the SAT in 1973 from 3 hours to 2 1/2 (in order to accommodate the half-hour Test of Standard Written English) might have had some relationship to the sharp drop that year in the SAT scores. We are satisfied that it did not. The most careful precautions possible were taken to assure that the reduced item selection would not result in changing the difficulty of the test.

We have also looked into the question of whether the decline in the SAT scores has affected their "validity" as predictors of individuals' college performance. It has not.

A detailed review made for the panel of the experience between 1964 and 1974 in a substantial number of colleges discloses some minor but interesting variations in the pattern of the SAT's predictive effectiveness, depending on the type of college involved (four-year or two-year), on the test takers' sex, and on whether they come from higher, lower, or middle SAT-score groupings. There were also some perhaps significant changes in this pattern during that 10-year period: the median validity coefficients (which are measures of the predictive capacity of the test) went down in general during the late 1960s and then went up sharply in the early 1970s. Some slight additional illumination is provided by looking at a number of studies that have been made of comparative validities of the SAT, as well as other academic predictors, for students with different ethnic backgrounds.

Although some of these variations in the validity pattern probably warrant further analysis for other reasons, they reveal no lessening of the value of the SAT as a predictor of first-year college academic accomplishment. The predictive validity of both the Verbal and the Mathematical parts of the SAT increased between 1970 and 1974 in the colleges that had validity studies made during that period, while the predictive validity of high school grade records was staying about level. High school grades are still the best single predictors of college performance, but when these grades are combined with SAT scores more accurate prediction proves possible. It illuminates this picture only for those expert in the field to note that, as of 1974, the median validity coefficients for the combined six samples used in the ETS study were .39 for the SAT-Mathematical score, .42 for the SAT-Verbal score, .50 for high school grade records, and .58 for the three predictors combined. The comparable median validity coefficients in 1970 were .29 for SAT-Mathematical, .37 for SAT-Verbal, .49

for high school grade records, .56 for the three predictors combined. For the layman, this says that the SAT remains a useful, but far from perfect, predictor of college performance.

- In general, and after checking the technical and psychometric aspects of the SAT thoroughly, the panel finds consistent confirmation that the score decline has not resulted from changes in the testing instrument. The scaling and equating and item-obsolescence procedures that are followed are reliable, and the predictive validity of the test is slightly higher than it was before. The standard established in this test has remained substantially constant, and the decline the scores reflect is, if anything, slightly larger than the reported record indicates.

A Changing Context?

The harder question to answer is about the possible effect on SAT scores of a changing "relevancy" of the test because of its being kept the same despite changes in curriculums, in educational policies and practices, and in learning's processes.

We consider later in the report (Part Four) whether the score decline may have resulted in part from reduced concentration on the "basics" of reading and writing and arithmetic and from a possible lowering of standards in elementary and secondary education. The "relevancy" point, however, though it is related in one sense to those questions, goes beyond them and has a different character. The suggestion is that in the 36 years since the present SAT standard was established the society has set new and different (as distinguished from higher or lower) learning goals, that the colleges or schools or both have adopted new priorities in education, that different learning and communication processes have come into use, and that part of the reason for the decline in the SAT averages is that the test does not reflect these changes.

In one view of this line of reasoning, there is a short and complete answer to it. If the value base of the SAT is accepted as being solely the prediction of college academic performance, the critical fact is that the test's predictive validity is actually somewhat higher than it used to be. The panel's further investigation reveals a similarly increasing positive correlation over the past 10 years between SAT scores and students' grades in high school. These two findings seem to leave little basis for the suggestion that the SAT has gotten out of line with either secondary or postsecondary practices or standards—so far, that is, as these practices and standards are reflected in students' academic grades.

This doesn't actually meet, however, the broader criticism of the college entrance examinations. It is partly an objection to the common misuse of the test scores. A student leader and valedictorian in a District of Columbia high school was recently denied admission to a college because his SAT scores were low. High schools are being measured by the averages the college-bound segment of these graduates get on the college entrance examinations. The College Board's warnings against these practices are disregarded. When these test scores and averages are used as exclusive or *overall* measurements of individual or institutional quality or accomplishment, a very real "relevancy" issue arises.

The challenge goes further. Recognizing the continuing correlation between high school grades, SAT scores, and college grades, the criticism is that this correlation covers less and less of what is important here. With grade inflation rampant in both secondary and postsecondary education, even while remedial courses have to be added constantly to first-year college curriculums, the argument seems increasingly pertinent. The correlation between grades and scores and then more grades could increase or decline or stay level, even while there was dangerous deterioration or healthy improvement *by broader measures* on both sides of the secondary/postsecondary divide. That correlation might come to apply to a smaller and smaller part of the learning actually going on.

This raises issues going beyond the panel's assignment and its competence. We have accepted, for purposes of this inquiry and report, the traditional value base of the SAT—its validity, that is, as an instrument helpful in the determination of students' likely academic performance. We have considered, in Part Four of the report, the question of whether less concentration on "basic" courses and an apparent relaxation of standards has had an effect on the SAT averages. So far as the relevancy issue goes beyond this it gets into broad and basic questions of educational policy.

Yet we feel strongly that this set of issues warrants further consideration by those more fully qualified than we are, working from a broader charter, for we have gotten a strong sense of possibly basic change taking place in the relationship between secondary and postsecondary education, and particularly in the function of testing in connection with the passage from high school to college.

We are concerned about the relationship between traditional testing techniques and the introduction into the learning processes of new forms of communication and teaching instruments: tape recorders, films and film strips, television (to which we return later), and the like. The lack of either quantifiable or other clear evidence is not enough answer to those who suggest that tests in the conventional college entrance examination form do not measure competencies developed in the schools through increasing reliance on new combinations of kinetic, audio, and visual—"kin-audio-visual"—teaching materials and processes.

We note the strongly held beliefs of many that different forms of testing would facilitate the effective and constructive transition of young people from high school to college. Questions are raised about the assumption that this testing should be based exclusively on predicting only college academic performance, particularly grades in first-year courses. There is the suggestion that consideration can and should be taken of the fact that different individuals are, as high school juniors and seniors, at such different stages of their personal development. Others question the effect of the "speededness" of tests, the use in them of uncommon words, and their possible penalization of test takers' putting ideas in personal terms.

A broader listing of questions of similar kind is included in the report of the *Conference on Declining Test Scores* held in June 1975 by the National Institute of Education. That report includes an invaluable summary of the areas in which additional research and analysis appear warranted.

The panel accordingly commends further inquiry by the Board and ETS into the function of tests at this critical passage point. In broader terms, the agencies administering college entrance examinations have a superior opportunity to engage in what is in effect a brokering function between secondary and postsecondary education in this country—a function we consider critically important and in general too little exercised by either public or private agencies.

We are not suggesting any compromising of the levels of acceptable educational standards, for we count those currently accepted too low. The purpose of such inquiry would be to identify and put in appropriate priority whatever can be distilled from current national concerns about the society's educational values, and then to evaluate the traditional tests in the light of that determination.

Part Three. The Two Score Declines

FOURTEEN years of uninterrupted decline in the SAT scores create the illusion that there is some single force or closely related set of forces at work here. This isn't the case. The decline has developed in two distinct stages, characterized by significantly different balances of materially different causal factors.

During the first six or seven years of the decline the composition of the SAT-taking population was changing markedly. Each year it included larger proportions of characteristically lower-scoring groups of students. This pulled the overall average down. There were only slight falloffs during that period in the score means *within* any particular ability groups.

The pattern changed after about 1970. The "compositional" shifts slowed down materially. What showed up increasingly was an across-the-board score decline, the apparent consequence of more "pervasive" changes or influences affecting higher- and lower-scoring groups alike.

Compositional Change

It is already hard to remember the extraordinary confluence of forces that struck the educational system in the 1960s.

About 1,864,000 students graduated from high school in 1960, and some 564,000 juniors and seniors took the SAT that year. Ten years later the number of graduates had increased by a million, and the number of SAT takers had tripled.

This was partly a demographic change. Between 1964 and 1965 the number of 18-year-olds in the country jumped by more than a million. That was when the post-World-War-II population wave first hit this age level. It was also the time the SAT average scores started down.

By what was probably more than coincidence, the nation decided during that same period to reduce the high school dropout rate and to see to it that a larger percentage of young people had the opportunity to go on to college.

Perhaps it was a historical accident that this was also the time of tardy legislative decision to attack previous discrimination in providing educational opportunity, particularly discrimination based on race, sex, and family income.

Twenty-five years ago, only half of all young Americans were staying in school through the 12th grade; this fraction grew by 1964 to two-thirds and by 1970 to three-fourths. The proportion going on to college was about one-fourth in 1952, about a third in 1964, and almost half by 1970. (There are different published school completion statistics. The panel has relied on the Census Bureau series, which measures retention in terms of the number of students who entered the fifth grade and are still in school at various subsequent grade levels.)

It would be pleasant to think that as increased percentages of vastly larger numbers of young people stay in school longer and go on to college, the college entrance examination averages achieved before by a favored fraction of students could be held constant. Yet any such expectation would be ruefully unrealistic. The major move toward equality of opportunity in the 1960s will be judged unfairly unless it is recognized that an increasing school retention rate is bound to mean, at least at first, some drop in the *average* developed ability level (as reflected in traditional tests) of the larger number staying the course. Yet the record for the years between 1950 and 1963 indicates that the effect of this need not be either serious or enduring.

The details of the changes that took place in the composition of the SAT population during the first period of the decline are hard to identify, for no biographical data were collected on these test takers before 1972. This kind of information is available, however, in connection with two other sets of tests: those given as part of Project TALENT in 1960 and those included in the National Longitudinal Study of the class of 1972. It has proved possible to equate the reading examinations on those two tests, to identify (from ETS records) students who took one or the other of them and also took the SAT, and so to reconstruct in terms of biographical data two SAT cohorts, one for 1960 and the other for 1972. There is also relevant evidence available from the Student Profile studies made since 1966 by the American College Testing (ACT) Program and from the series of Freshman National Norm studies made by the American Council on Education (ACE).

What is clearest from all this data is that, starting in about the mid-1960s, cumulatively larger *percentages* of students with comparatively lower high school grade averages were going on to college. The ACE Freshman National Norm study shows this directly, as do the ACT Student Profile data for those students taking that college entrance examination.

This shift appears to have been sharpest of all among the SAT-taking group. The composite study (of Project TALENT, National Longitudinal Study, and SAT data) shows this in various ways, based on a comparison of scaled score averages on the equated 1960 and 1972 reading tests of all these groups. The averages of the SAT takers dropped by twice as much on these reading tests as did those of high school seniors as a whole, reflecting the increasing percentage of lower scorers among those taking the SAT. In 1960, over half (55.4 percent) of the SAT takers came from the highest-scoring groups (top 20 percent) on the Project TALENT and National Longitudinal Study reading tests; in 1972, this had dropped to a little over a third (36.4 percent). The significance of this decline among the SAT takers is emphasized by the fact that a 1970 repeat of the 1960 Project TALENT Reading Comprehension Test showed a slight gain among 11th-grade students *as a whole*—despite the fact that the proportion of the age group staying on in school through that grade level had increased from 77 percent in 1960 to 87 percent in 1970.

There is limited usefulness, however, in a finding that the first stage of the SAT score decline resulted in large part from the fact that proportionately more students with demonstrated lower scoring records began taking it. This advances the analysis of what happened here only from one set of examination scores to another. To look with constructive intent for the root causes of the decline requires pressing further into the *reasons* for these differences in scoring capacity.

There were significant increases during the earlier period of the SAT score decline in the proportionate numbers in the test-taking population of three groups that have always registered substantially lower-than-average scores on this test: students from lower-socioeconomic-status families, members of minority ethnic groups, and (on the Mathematical but not on the Verbal portion of the test) women.

Students from families with the lowest incomes (under $6,000 in 1977 figures) average about 100 points lower on both the Verbal and Mathematical parts of the SAT than do those from families with the highest incomes ($18,000 and over); and the score averages go up consistently from one intermediate income level to the next. Students' scores also vary, on the average though with many exceptions, with the educational attainment levels of their parents.

An analysis of the 1960 Project TALENT and the 1972 National Longitudinal Study data shows substantial increases over that period in the percentages of students from the lower socioeconomic quarters who went on to college and decreases (although these characteristics are not direct functions of each other) in the percentages from the two higher quarters. These changes are more marked than those emerging when comparisons are made in terms of *ability* (as identified in terms of high school grades). When these figures are broken down to separate students going on to four-year colleges (principal users of the SAT) from those entering two-year colleges, the shifts become even larger with respect to the four-year college group.

Although the data regarding ethnic group score differentials and test-taking population changes during the 1960s are sparse, the picture is relatively clear with respect to whites and blacks (but with too little information to go on, as far as other minority groups are concerned).

Very few blacks or other minority ethnic group members were taking the SAT in the early 1960s. As of 1966, some 5 percent of the first-year college students reported on the ACE National Norms Study were black (and 91 percent white), but few of them were in colleges using the SAT. An estimate of black SAT takers in 1963 of between 1 and 2 percent of the SAT population would probably be high. The ACE studies show the percentage of black freshmen rising to 8.7 percent by 1972 and remaining at about that level since that time. By that year, too, the SAT Student Descriptive Questionnaire showed about 8 percent of the test takers to be black, and this figure has gone up only slightly in subsequent years.

A 1965-66 study of the equality of educational opportunity, made by James S. Coleman for the United States Commissioner of Education, reports 12th-grade-level achievement score averages for blacks "about one standard deviation below those of . . . whites, which means that about 85 percent of the [black] scores are below the white average." A roughly comparable picture is provided by the Student Descriptive Questionnaire data, with blacks averaging approximately 100 points below the overall average on the Verbal and about 115 points lower on the Mathematical part of the SAT.

Women and men have traditionally averaged about the same scores on the Verbal portion of the SAT, but there has been a marked difference in the Mathematical averages. In 1960 the Mathematical means (as derived from the composite study made for the panel) were 465 for women, 520 for men. Twelve years later, the average for women was virtually unchanged, but the average for men had dropped by 14 points (to 506). The 1977 Mathematical figures are 445 for women, 497 for men.

Women represented 42.7 percent of the SAT-taking group in 1960 and 47.5 percent in 1970.

Yet such groupings of SAT takers, like those in terms of grades on other tests, relate to only the superficial aspects of the score decline. The *causes* of whatever is reflected here lie in the *reasons* these groups score lower.

The suggestion is sometimes made that the SAT is culturally biased. Definitive analysis of cultural bias is virtually impossible. These same differences show up on most other standardized tests, and yet this proves nothing. The panel's inquiry into the test design procedures followed by ETS confirms that special efforts have been made to avoid the suggested prejudices. Cultural bias would appear to be more likely to affect the Verbal part than it would the Mathematical part of the test; but the differences between the averages for various ethnic groups are larger for Mathematical scores than they are for Verbal scores. Although the available information is incomplete, the predictive validity of the SAT *appears* to be substantially the same for students in different ethnic groups and for women and men.

The significant "biases" involved here clearly go much deeper and concern the society more than the tests.

We struggle as a nation against the fact that the unevenness of children's educational accomplishments parallels, in general though with many exceptions, the educational and economic attainments of their parents. When an increase in the percentage of economically disadvantaged students staying in high school and taking college entrance examinations results in lowering the average scores on these tests, one thing this says is that the national effort to neutralize this kind of disadvantage is still incomplete.

Score differences between blacks and whites parallel closely the differences in averages between students from low- and high-income families and between those whose parents have differing levels of education. Beyond this, two centuries of racial bigotry have unquestionably left an educational system that still serves blacks and other minority groups less well than whites, particularly when it comes to meeting traditionally accepted "majority" standards. The contributing cause of the score decline is not that more minority group members now take the SAT, but that despite statutory guarantees of equal opportunity the society has not yet developed either the educational means or the mores that will bring children with different racial roots to a parity of aptitude—as the SAT and other tests measure it—by the time they reach the 12th grade.

That women score lower than men on the Mathematical sections of the SAT almost unquestionably reflects more than anything else the traditional sex stereotyping of career opportunities and expectations.

In all these respects, the national decision made in the 1960s was to attempt to neutralize the discriminatory aspects of cultural differentiations. To the extent that less-economically advantaged young people, minority group members, and women still score lower on college entrance examinations than do their counterparts, one of the causes of that score decline is properly identified as being the incompleteness so far of the society's now-earnest effort to be honest with its expressed ideals.

Realizing that even recognition of these group differentials risks irresponsible headlines, attributing the score decline to more blacks' and women's and poorer youths' taking the test, we note the figures. Even if blacks' lower scores were identified solely with race (which would ignore other factors involved here, such as income levels), their increased taking of the SAT would account for no more than 4 or 5 points of the total decline in averages over the entire 14-year period. Women's larger participation could be identified with about the same amount of the drop in the Mathematical average, but with none of the decline in Verbal scores. The effect of increased test taking by students from lower-income families cannot be this specifically quantified; it is larger. The figures for all three groups overlap. There is no legitimate short answer to be found in this set of developments alone.

Another tempting shortcut to an answer that perhaps appeals to quite a few is that there are simply too many young people now going on to college. We do not find the support for that conclusion in this evidence. Should the next 15 years see the opening of still further postsecondary educational opportunities to young people who today are not finishing high school, with another attendant decline in test scores, we suspect that the reaction would again be to welcome the problem on the one hand and on the other to recommend changes in the schools in order to meet it.

The panel has looked extensively into what is in effect another dimension of this compositional shift in the SAT-taking population, involving changing college-going patterns.

Fifteen years ago, most SAT takers were students enroute to relatively prestigious and selective four-year, liberal arts colleges and universities. While the *number* of those in the SAT population going on to this type of postsecondary institutions remained virtually constant at first (until 1967), a cumulatively increasing *percentage* of test takers began to follow different courses: to colleges and universities with less selective or even open admission policies, to two-year colleges, to training with a more technical or vocational emphasis. Most of this change, though not all of it, had taken place before 1970 or 1971.

There are average score differentials of from 60 to 85 points (1) between test takers going on to four-year colleges (who average higher scores) and those who subsequently enter two-year colleges, and (2) between test takers who go directly from high school to college (averaging higher scores) and those who do not. There were substantially larger percentages of these two lower-scoring groups in the SAT population in 1972 than there had been in 1960; the increase in both cases was from about 8 percent (in 1960) to approximately 15 percent (in 1972).

There were comparable *drops* during that same period in the percentages of SAT takers asking to have their scores sent to highly selective liberal arts colleges (from 13.2 percent in 1961 to 5.6 percent in 1974) and to research universities (from 10.5

percent to 6.4 percent); and these are groups who have traditionally scored *higher* on the SAT.

These shifts in college-going patterns were more than just reflections of the compositional changes already identified. It was during the 1960s and early 1970s that two-year community colleges grew so fast and that more flexible admissions policies were adopted by many four-year colleges and universities. The SAT, originally used as an instrument for assessing the abilities of a comparatively small group of high school students to do a particular type of postsecondary work, came to be taken increasingly by a much wider variety of students with more diverse prospects in mind.

One aspect of this change in college entrance policies was a marked diminution in the percentage of students taking the SAT more than once. Since repeaters average 15 to 30 points higher the second time they take the test (whether because of the repeating or for other reasons), the significant decline in the number of repeaters has had some small effect on the score averages. We have not broken this down into pre-1970 and subsequent years.

There is a question of whether "cramming" for the SAT is a factor here. The College Board and ETS have taken the general position, based in part on studies that may now be out of date, that this type of preparation for these tests is not effective. Although we question this, we have not pursued the point, partly because this can in no event be a significant element in the score decline, and partly because the sponsoring agencies have now taken this question under active review.

■ We find, therefore, that the largest part of the SAT score decline between 1963 and about 1970 was identifiable with compositional changes in the mix of the SAT-taking group—considered both in terms of the test takers coming from higher- and lower-scoring groups and in terms of their plans for going on to college. Although precise identification of the degree to which these changes explain that part of the decline is impossible, fairly careful calculation indicates that they account for between two-thirds and three-fourths of it.

Already appearing during that period, however, were indications of a broader set of influences on these scores—which were subsequently to emerge more plainly and strongly.

Pervasive Change

Neither reason nor data would suggest that at some precise time the changes in the test-taking population, which had been moving the SAT averages down, ended abruptly and that a new set of forces then erupted suddenly to continue and even accelerate the previous decline. This did not happen. But under this heading we want to develop the concept that after 1970 changes in the test-taking group became less important as a cause of the score decline and other factors in the schools and in the society at large became more significant.

By 1970, the previous increases in the percentage of young people finishing high school, going on to college, and taking the SAT had either stopped or had started to reverse. It is also clear that there has been much less shifting in the past several years in the proportions of characteristically higher- and lower-scoring groups of SAT takers, as compared to the 1963-70 period.

A relatively complete picture of the test-taking population is available from 1972 on, from the Student Descriptive Questionnaire (SDQ), which is filled out by more than 80 percent of all SAT takers and which covers some 70 informational items. A correlation of SAT scores and SDQ answers permits a year-by-year comparison of the average scores of more than 200 groups and subgroups of test takers, along with an identification of the proportion each group represents of the total SAT-taking population. The SDQ was first used in 1971-72, but was changed so materially the following year that the 1972-73 information constitutes the earliest base for reliable comparison with subsequent years. It appears, however, from other indexes and studies that most elements in the compositional pattern reflected in the SDQ data had taken shape from one to three years before that questionnaire was fully developed. The precise time factor is not significant.

The SDQ data show a limited degree of continuing compositional change in the test-taking group after 1972. The percentage of women taking the SAT has increased from 48.7 percent in 1972, to 51.1 percent in 1977—enough to account for several points of the decline in Mathematical scores. There has also been some continuing increase in the number and percentage of test takers from minority ethnic groups, although it appears to be small.

One set of SDQ questions shows that substantially more students now than in 1973 indicate an intention to pursue "occupational" or "career" majors, with proportionately fewer apparently expecting to go on to "arts and sciences" majors; and those in the former of these groups show consistently and markedly lower average scores on the SAT (Verbal and Mathematical sections alike) than do those indicating the "arts and sciences" choices. We have looked into this situation thoroughly, taking account of a possibly significant change in the form of the SDQ question, of answers to other apparently related questions, and also of the reports regarding similar questioning done on the Project TALENT and the ACE Freshman National Norms studies. While the particular point is probably of secondary importance, it confirms the broader fact that there continues to be some "compositional" change taking place within the SAT population.

The SDQ data show only insignificant changes between 1973 and 1977 in the distribution of SAT takers among students at various high school grade percentiles. It should be noted here, and more generally as well, that the SDQ data are all based on students' own answers to these questions and may therefore reflect certain biases. There have also been changes in the form of some questions, including the one about rank in high school classes, which may affect the year-to-year comparisons.

Some of the information initially submitted to the panel suggested an aberrational drop since 1970 in the number of students scoring 600 or above on either the Verbal or the Mathematical sections of the test or on both. That "high scorer" figure, which stood at 189,300 in 1970 (among *all* SAT takers, as distinguished from a one-year cohort), had gone down by 1976 to 108,200; and this drop seemed to suggest elements of compositional or other related change.

An inordinately extended analysis of the drop in the number of these "high-scorers" indicates, however, that the decline probably results almost entirely from the reduction in the number of students taking the SAT and from the impact at the top of the same pervasive influences that have been affecting the scores of the test takers as a whole and which we discuss in the remainder of the report. This was confirmed by a special study the panel had made of the SAT scores of some 1,500 valedictorians and salutatorians in 145 high schools during the period between 1960 and 1974. The study shows that there were no significant changes in the average scores of these top-ranking students during the early period of the overall SAT score decline, but that they dropped after 1970 on the Verbal part and after 1968 on the Mathematical part of the test at about the same rate of decline in the overall averages – making these data particularly direct evidence of the panel's "two-decline" conclusion.

In general, comparatively little – perhaps 20 to 30 percent – of the SAT score decline during the past five to seven years appears attributable to compositional shifts within the test-taking group.

There is consistent evidence, at the same time, of the emergence during this more recent period of what the panel has characterized as "pervasive" score decline – in the sense that it has shown up within virtually all categories of SAT takers.

Score averages have gone down since 1973 (with the drop apparently having started two or three years before that) among students at the higher and those at the lower percentiles of their high school classes, among students in private and in public schools, among those in large and in small high schools, among those taking "academic" and those taking "career" courses of study in high school, among test takers from high- and from low-income families, among men and women, among white students and those from minority groups, among students expecting to go on to different kinds of colleges, among those intending to take postgraduate work and those looking only toward a baccalaureate.

Although the drops in score averages have varied somewhat in degree among different groups, the obviously significant characteristic of the pattern in this second stage of the SAT decline has been the pervasiveness of the effect of the forces influencing it. The statistics themselves say nothing, at least directly, of the nature of these forces. We have set this question apart for separate consideration in Part Four.

We note, although it is too early to appraise its significance, that the SAT scores have changed less in the past two years. The 1977 high school seniors' Verbal average is 5 points below what it was in 1975, and the Mathematical average is down only 2 points. The drops during the preceding two-year period had been 11 points on the Verbal part and 9 on the Mathematical.

If the figures for all candidates are taken, which means particularly including high school juniors, the 1977 Verbal average is the same as the average last year, and the Mathematical average has gone up 1 point.

Through Other Looking Glasses

To establish further basis for determining the implications of the apparent two-stage pattern of the SAT score decline and also to check this analysis, the panel has reviewed what has been happening during recent years on other standardized academic tests. Several of these are related in one way or another to the SAT itself.

We find, in general, that the SAT score decline is significantly consistent with a broader pattern. Most—though not all—standardized test score averages, which had previously been rising gradually, turned and started down in the middle 1960s. In virtually all instances there has been a much sharper drop in the 1970s. Most of what appear at first to be inconsistencies between the records on the various sets of tests became reconcilable when their histories are analyzed on the "two-decline" basis. In significant respects, at the same time, the comparison with other test score experience provides an important reminder that declining college entrance examination scores do not in themselves warrant generalization about what is happening to the abilities at large of youth as a whole.

The other widely used college entrance examination, administered by the American College Testing (ACT) Program, shows a comparable decline since the mid-1960s, with most of the drop having taken place since 1970. The averages have dropped more on the Mathematics section of the ACT than on the English section. That test also includes a Natural Science component, on which there has been no decline, and one in Social Science, which shows the largest drop. Women have consistently scored lower than men on the ACT Mathematics test but higher on the English component.

American College Testing Program representatives advise the panel that the ACT composite score average will be slightly higher for 1977 than it was last year. Taken with the current experience on the SAT scores, this affords an interesting basis for conjecture as to whether what is involved here is a pause or a possible turning point.

Score averages on tests taken by high school seniors as a whole, as distinguished from those taken only by students going on to college, remained virtually level until about 1970; but where more recent reports are available (as, for example, on the Iowa Tests of Educational Development and the Minnesota Scholastic Aptitude Test) they show declines in the last five years roughly paralleling those on the SAT and the ACT.

Score declines have been largest at the 12th-grade level and somewhat less at each successively lower grade, with the averages at grades 1 to 4 remaining relatively constant. The panel has not been in a position to analyze the extent to which this reflects various possible explanations: decreasing school system effectiveness at the upper elementary and secondary levels, increasing school retention rates, changing motivational factors, possibly diminishing relevance of some of the standardized tests to what is being taught.

We have given particular attention to the results of two other sets of examinations sponsored and administered by the College Board and ETS. One of these is the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test, a shorter version of the SAT taken annually by over a million high school students, most of them in their junior year. (The test has also been used since 1971-72 by its cosponsor, the National Merit Scholarship Corporation, as a qualifying test.)

For reasons and with implications that are only partially clear, the PSAT/NMSQT scores followed a different pattern than did those on the SAT between 1963 and 1973. The declines were substantially smaller. Furthermore, when the PSAT/NMSQT was given in 1960, again in 1966, and once more in 1974, to national samples of 11th graders as a whole (not being limited, therefore, to students probably college bound), these "norming studies" showed a substantial stability in averages on both the Verbal and Mathematical sections over the entire 14-year period.

An accounting of the complex variety of factors that emerged in the panel's examination of this difference between the SAT and PSAT/NMSQT patterns would burden this report unduly. We find adequate explanation for most of the difference but not for all of it. There may be support here for the view of some analysts that there has been more change taking place at the 12th- than at the 11th-grade level. What emerges as most significant, however, is that since 1973, both the Verbal and the Mathematical score averages on the PSAT/NMSQT have dropped in almost exact parallels to the declines on the SAT. So here again, the "two decline" pattern shows up, though with a perhaps noteworthy difference in timing.

There has been a probably significant difference in the pattern of scores on the Achievement Tests, covering 14 different subject-matter areas, which some high school juniors and seniors take in conjunction with the SAT. Some colleges and universities require applicants for admission to take certain of these tests; others recommend it. Students may themselves choose to take various of the Achievement Tests and to have the scores on them transmitted along with their SAT score. Each of two of these tests (in English Composition and Mathematics Level I) is taken by more than 175,000 students (high school juniors and seniors) each year; each of eight others (American History and Social Studies, French, Spanish, Biology, Chemistry, Physics, Mathematics Level II, and Literature) is taken by between 15,000 and 75,000 students; fewer than 10,000 students take the German, Hebrew, Latin, Russian, and European History tests. Use of the Achievement Tests has declined sharply in the past four or five years. Students taking them are characteristically of higher-than-average ability.

Although the group taking Achievement Tests is a special and relatively small one, several elements in the picture nevertheless appear important. There have been relatively small declines in the past 10 years on four of the more widely used Achievement Tests (American History and Social Studies, Mathematics I and II, and Literature) and *increases* on the six others (English Composition, French, Spanish, Biology, Chemistry, and Physics), during what was the period of most of the decline on the SAT scores themselves.

When students' scores on the Achievement Tests are compared with their scores on the SAT, it is apparent that in those six Achievement Test groups for which scores on the Achievement Tests went up between 1967 and 1976, the students' SAT-Verbal score averages went down. This was not true, however, as far as the SAT-Mathematical scores were concerned; all six groups showing increases in Achievement Test score averages also showed increases over the nine-year period in SAT-Mathematical averages.

We have not been able to analyze fully the possible implications of this apparent divergence between aptitude and achievement test score patterns. Perhaps part of the explanation is that the students taking the Achievement Tests are particularly highly motivated, although this does not explain the different patterns of their performance on the Verbal and Mathematical sections of the SAT. It is conceivably important that the College Board and ETS make much larger use of outside committees in connection with the Achievement Tests than with the SAT; the counsel sought is in the one case from experts in the particular disciplines, in the other more from psychometricians and psychologists. There are possibly clues here to significant questions regarding the changing "relevance" of one set of tests or another. This body of data warrants further analysis by the Board and ETS.

A mixed pattern of developments emerges in the scores on tests that college students take enroute to postgraduate study. The ETS Graduate Record Examinations averages (on the Verbal and Quantitative sections of that test) have declined at about the same rate since 1967 as have the SAT-Verbal and Mathematical averages, with almost half the drop concentrated in 1969-70. There is a similar pattern on the Graduate Management Admission Test. The Law School Admission Test and the Medical College Admission Test show increases, however, apparently reflecting primarily the increasing competition for graduate school admission in these two areas.

Only preliminary results are available so far from the National Assessment of Educational Progress (NAEP), a project that will eventually provide tests at four-year intervals of national samples of individuals at various age levels (9, 13, 17, and 26-35) in 10 different curricular areas, with a considerable amount of accompanying biographical data being supplied by the test takers. Of the four banks of NAEP tests completed so far, three are in the Reading and Writing areas. They show results differing considerably from those on most other standardized examinations. The NAEP tests in Reading Performance and in Writing Mechanics show relatively slight declines in the averages for the 17-year-old group between 1969-70 and 1973, but the test in Functional Literacy and Basic Reading Performance shows a small *increase*—particularly among the groups of students (blacks, for example, and children from lower socioeconomic families) who scored lowest on the first round of tests. NAEP Reading Performance and Writing Mechanics tests given to the 9-year-old and 13-year-old groups show the usual pattern of lesser declines (or even increases) at the lower age levels.

Although it is proposed to extend the NAEP testing project eventually to include not only enrolled students but those out of school (in the 17 and 26-35-year-old groups), very little has yet been done. So little is known about the performance of age groups as a whole on tests of this kind, except for the familiar evidence of the results, while the military draft was in effect, of the Armed Forces Qualification Test (AFQT). The average on that preinduction test went up sharply between the mid-1960s and the early 1970s. There are, however, so many possible explanations of this rise, which came during the period when most tests of *students in school* were showing decline in this age group, that comparative analysis is exceedingly difficult. There are no reliable comprehensive measures yet of the comparative competence of today's youth with yesterday's. It could occur at the same time (1) that a larger percentage of young people going on to college would be less well equipped for what college has traditionally required, while (2) the general ability level of youth as a whole increased.

- In summary, the evidence regarding the two stages or elements of the SAT score decline is illuminated by the data regarding what has been happening on other standardized academic tests.

Although the drop in the SAT averages (especially the Verbal scores) has been larger than that on any of the other widely used examinations, it is consistent with a general pattern of decline found in other test scores, which started to develop in the mid-1960s and has persisted ever since.

A comparative analysis of these various test score records confirms, at the same time, the necessity of recognizing clearly the dual developments this pattern reflects. First, a larger percentage of young people are staying in school through the 12th grade with the consequence that there has been a lowering of the *averages* on examinations taken previously by more select groups. Second, it has emerged as a major factor in the more recent stages of these test score patterns that there are broader influences operating to reduce young people's developed ability levels, at least at the upper high school level.

The panel has identified, with relatively firm statistical basis, roughly two-thirds to three-fourths of the decline between 1963 and about 1970 as resulting from "compositional" changes in the SAT-taking population. As nearly as we can tell, though the statistics are less helpful here, only about a quarter of the subsequent decline is traceable to continuing compositional change, with the remaining three quarters apparently resulting from the impact of more "pervasive" forces.

We turn now to what analysis is possible of the nature of this second set of influences.

Part Four. Circumstantial Evidence

Searching for the causes of the SAT score decline over the past six or seven years is essentially an exercise in conjecture. So much has happened that may have affected this record that there is no way of telling what did; the only evidence is circumstantial, leaving it hard to distinguish cause from coincidence. Most of the 50 or so theories brought to the panel's attention have in common only three assumptions; first, that since the problem has been reduced to a single statistic – the drop in these averages – there must be a single answer; second, that what has happened is in every respect bad; and third, that whatever caused it is somebody else's fault.

Although the panel's only certain conclusion is that we are dealing here with a virtually seamless web of causal connections, the apparently most critical elements emerge more clearly in looking first at some developments in the schools, then at several major societal changes, and finally at the murky but probably vital area of youths' motivations.

In the Schools

While some of the observable changes in formal education during the past 20 to 25 years probably reflect evolving educational philosophy, our impression is that most of them are traceable in larger degree to the vast expansion and extension of educational opportunity in the 1950s and 1960s. The *direct* effects of "compositional" change on the SAT score averages were concentrated in the period before 1970, but this leaves the question of how much fallout effect those changes have had since that time. There is the additional question of the relationship between what has been happening in the schools and what has taken place in learning's broader societal context. Looking for causes of the SAT score decline requires going beyond the evidence of revisions in traditional educational practices to the reasons these revisions have been made.

Courses of Study

A good deal of attention has been focused on the fact that fewer "basic" courses are now being required of all students in high school, with many more "electives" being introduced into the curriculum. This is asserted to be particularly true in the English and verbal skills area, and the evidence suggests that it is.

Using nationwide data assembled by the National Center for Education Statistics, Harnischfeger and Wiley report that in one recent two-year period (1971 to 1973) English enrollments in grades 7 through 12 dropped almost 11 percent, about 50 percent in Advanced English. This drop probably involves less those students who will be taking the SAT than it does the high school student population as a whole; the Student Descriptive Questionnaire shows about 90 percent of the SAT takers reporting consistently from year to year that they have taken four "English" courses. Yet the question remains as to what kinds of courses these are.

A study by the Massachusetts Department of Education combines a survey of the development of "elective" courses in Massachusetts high schools between 1971 and 1976 with an analysis of the SAT scores in 43 of these schools—divided into three groups on the basis of how their students' SAT averages compare (in terms of change between 1971 and 1976) with the national SAT averages.

During the five-year period there were increases of over 50 percent in the number of English/Language Arts courses offered (at least as indicated by course titles) in the 43 Massachusetts high schools; the two most commonly added were Science Fiction and Radio/Television/Film. No significant correlation emerges between the number of electives added by a particular high school and the experience of students from that school on the SAT. Yet when the record of student enrollments in the two commonest "specialty" courses is analyzed, it develops with significant consistency that those schools that showed increases in these enrollments between 1971 and 1976 also showed larger than normal declines in SAT scores; and where the specialty course enrollments went down or stayed about the same, there were not substantial SAT score declines. Another Massachusetts study shows that between 1968 and 1973 over a quarter of all high schools in that State added courses in Film Making; the number offering 11th-grade English and World History courses went down.

There have been similar developments in other states. In California, for example, it is reported that "enrollment in basic English courses fell 19 percent between 1971-72 and 1974-75 and in English Composition classes it plummeted 77 percent. . . . Enrollment in contemporary literature electives (Children's Theater, Mystery and Detective Story, Executive English) has nearly doubled."

While the panel's net conclusion is that there is almost certainly some causal relationship between the shift in the high schools from courses in the traditional disciplines to newer electives and the decline in SAT-Verbal scores, we warn against any oversimplistic interpretation of this finding.

A careful probing of the evidence indicates that the new electives are being taken less by students who are going on to college (and will therefore take the SAT) than by those who are not. It will have to be determined whether the needs and interests and developed competencies *of those taking these electives* are better met by a course, for example, in Radio/Television/Film or an English IV course in the refinements of the language. Nor has it been established that even with respect to SAT takers these elective courses contribute *by their nature* less to the "aptitudes" the SAT supposedly measures—distinguishing this, to the extent such a distinction is possible, from "achievement." In our view, "returning to the basics" would be wrong unless it included full reappraisal of what the right basics are—taking account of childrens' different rates and modes of learning and their different interests and plans for the future. The need is not to revert to uniform drills and exercises commended only by a traditional pedagogy, but to move ahead to a larger emphasis on the fundamentals of learning that can be identified as strengthening the base on which all students can build.

Our firmest conclusion is that the critical factors in the relationship between curricular change and the SAT scores are (1) that less thoughtful and critical reading is now being demanded and done, and (2) that careful writing has apparently about gone out of style. So we do not identify the score decline narrowly with reduced high school offerings of whatever used to be included in Advanced English courses. There is as much opportunity, and sometimes more incentive, for worthwhile reading and responsible writing in subject-matter areas of particular student interest.

We can't prove that learning how to write is related to a decline in scores on a test that requires no writing. Yet in our judgment this may be a significant factor. We suspect strongly that *expressing* something clearly and correctly—especially in writing—is thinking's sternest discipline. A recent study by National Assessment of Educational Progress of the number of writing assignments given a group of 11th-grade students shows that during a six-week period more than half of them were asked to write three or fewer papers, 12 percent only one, 13 percent none.

It seems clear that increasing reliance in colleges and high schools on tests requiring only the putting of X's in boxes contributes to juvenile writing delinquency. Students learn what they think they need to know. There is more than irony in the report of teachers, who used to train students to write, now advising them about the advantage in using soft-tip pens and pencils so these boxes can be filled in more quickly.

We applaud the recently announced reintroduction this year, at the December test administration only, of a written essay as part of the College Board's English Composition Achievement Test.

Our strong conviction is that concern about declining SAT-Verbal scores can profitably be concentrated on seeing to it that young people do more reading that enhances vocabulary and enlarges knowledge and experience, and more writing that makes fledgling ideas test and strengthen their wings.

It has been pressed strongly on the panel that there is a related development in the sharp decline in enrollments in high school foreign language courses, ancient and modern. Such courses, it is suggested, increase the knowledge and understanding of the English language—a development reflected in performance on the Verbal part of the SAT.

A clear parallel unquestionably shows up between students' SAT-Verbal scores and the number of foreign language courses they have taken in high school. Those who report having taken four or more such courses (about 10 percent of the test takers) average more than 100 points higher than those (about 8 percent) reporting no work in foreign language; and the averages rise progressively with the number of courses taken.

Is this, though, a *causal* relationship? Does foreign language training improve verbal ability or do students with high verbal ability take more foreign language? Is there anything here except confirmation that abler and more highly motivated students are more likely than others to take some foreign language course? It turns out that the foreign language students do as much better than others on the Mathematical part of the SAT as they do on the Verbal. The Student Descriptive Questionnaire reports show no reduction since 1973 in the amount of high school foreign language being taken *by students who also take the SAT*. Furthermore, although students who have taken foreign language continue to average higher SAT scores, the *decline* in the averages over the past four years has been at least as large as the decline among students with no foreign language exposure.

The panel may have given too little consideration to the difference between the score decline patterns on the Verbal and the Mathematical sections of the SAT. Most of the deviation had developed before 1963; the Verbal score average was at that time already 24 points below the Mathematical average. The two averages dropped in virtual parallel between 1964 and 1970. This gap has widened significantly, however, during the later stage of the SAT decline, until it is now 41 points. At one large and prestigious Midwestern university the median SAT-Verbal score of entering first-year students dropped by 60 points (from 580 to 520) between 1970 and 1976, while the Mathematical median score went down only 20 points (from 620 to 600) – with the high school grade averages of the entering classes remaining at a virtually constant level.

There may well be clues in this difference to some of the causal elements in the general score decline.

It is perhaps relevant that there has not been the proliferation of electives in high school Mathematics that has been characteristic of the verbal skills area. While Massachusetts high schools were increasing their course offerings in English by 50 percent (between 1971 and 1976) there was virtually no increase in the number of Mathematics courses being offered. Student enrollments in high school Mathematics courses have gone down some, but markedly less than enrollments in English courses.

We suspect, too, that there is more evidence here of the importance to SAT-Verbal scores of the diminishing emphasis on reading and writing, for these skills play much less part in mathematics.

The implications of the difference in scores on the two parts of the SAT may go considerably beyond this. We have conjectured that mathematics is essentially school-based learning, while verbal skills are more influenced by experience in the home.

Why, referring back to the point covered earlier, do students taking the College Board Achievement Tests, and maintaining previous scoring levels on them, show declining SAT-Verbal averages but stable or even increasing SAT-Mathematical averages?

The realization that the decline in SAT-Mathematical averages has been at about the same rate as the decline on most other standardized academic tests, while the SAT-Verbal decline has been substantially larger than any of the others, obviously raises the question – though we imply nothing regarding an answer – of whether there is something here that is peculiar to the Verbal part of the test.

Not having gone into the matter of the difference between these two score patterns as fully as hindsight suggests might have been advisable, we commend further consideration and investigation by the College Board and ETS.

Learning Standards

There have unquestionably been changes over the past 10 to 15 years in the standards to which students at all levels of education are held. Absenteeism formerly considered intolerable is now condoned. An "A" or "B" means a good deal less than it used to. Promotion from one grade to another has become almost automatic. Homework has apparently been cut about in half. Open admissions colleges are available; if entering students don't know how to read and write and do arithmetic, "remediation" is available.

There can be little doubt, despite a dearth of direct data, that these changes in standards are reflected in the decline in SAT score averages. The harder questions are about the reasons behind these changes, and about how to meet the problems presented here without compromising some interests and values and principles that the college entrance examinations do not pretend to take into account.

Several of these developments warrant particular attention.

Although the daily attendance and absenteeism data are infected by administrative chagrin attending their reporting, the available figures reveal a seriously worsening situation in the high schools during the late 1960s and early 1970s. Average daily attendance rates went down as the school retention rates went up. A 1975 report by the National Association of Secondary-School Principals identified absenteeism as "the most perplexing student problem. . . ." Absenteeism rates above 15 percent became common and 20 to 25 percent not unusual. There has apparently been some recent improvement in this situation.

Excessive student absenteeism has a doubly corrosive effect: on the absentee and on the class as a whole, when so many of those present don't know what was covered yesterday that the teacher repeats the lesson in an attempt to bring everybody up to pace.

The schools are both contributing cause and victim of this phenomenon. To the extent school is a place youngsters flee because they find themselves diminished or bored or both, the need is plainly for internal reform. Yet without the right alliance with home and community the school's effectiveness is limited. We note the probable need here for improved guidance and attendance services – which cost money few school budgets provide for this purpose.

"Grade inflation" is perhaps most commonly pointed to as a reflection of declining educational standards. An American College Testing Program study shows an increase, between 1962-65 and 1974-75, of 25 percent in the proportion of "A" and "B" grades reported by college-bound seniors in high school English courses. Other reports reveal a comparable pattern throughout both secondary and postsecondary education.

Yet we question the significance of grade inflation *as such* so far as the decline in college entrance examination scores is concerned. While the Student Descriptive Questionnaire data clearly confirm the fact that more students are now getting A's and B's than used to be the case, the distribution of SAT takers among the various *percentiles* on high school grade records apparently has not changed substantially.

The apparently more significant development involves the increasing acceptance of the notion that advancement from one grade level to another is an entitlement rather than something to be earned – or denied. It is unlikely that very many students who are promoted when their work doesn't warrant it will ever take the college entrance examinations. Yet here again, the necessary assumption is that other students' education suffers from teachers' having to gear their instruction to classes including too many who have fallen badly behind the rest.

The panel considers this, however, a hard problem with no easy answer. Automatic promotion hasn't developed because either teachers or principals like the idea. They don't. They know, however, the dubious value there is in "holding a student back" — so far as both that student and the rest of the class are concerned.

Now there is a nationwide movement, citing what has happened on the SAT and other comparable tests to support it, toward conditioning students' promotion at various grade levels on their achievement of prescribed scores on standardized tests. Reports, principally in the press, are that this is having a salutary effect. Yet dependence on rigid cutting points on test scores for this purpose makes no more sense than it does for college admissions. Perhaps something of this kind is valuable in reestablishing a commitment among teachers, students, parents, and community to the ideal of educational excellence. We suspect, however, that the promotion of students from one grade level to the next ought to be ultimately a matter of case-by-case decision, taking account of each student's social and psychological development as well as his or her intellectual growth and permitting flexible arrangements for moving ahead, for example, in certain subject-matter areas but not in others.

We conclude, in short, that what is reflected in the practice of automatic promotions has indeed been one of the causes of the SAT score decline, but we question an equally "automatic" answer that such promotion should depend entirely on scores on still other standardized tests.

The panel has considered, too, the relatively clear evidence (though none of it has been quantified) of reduced assignments of homework. We find this another reflection of the changing standards of elementary and secondary education, which seems to us related to the SAT score decline. There is no way of determining whether the reason for the reduction in homework was that teachers or that parents decided it was less important. We assume that television is a factor here, and we come to that problem shortly.

Trying to probe beyond these more obvious reflections of changed educational standards, the panel commissioned a preliminary investigation of the possible relationship between the SAT score decline and whatever changes there may have been in the textbooks being used at various primary, elementary, and secondary levels. Taking the Verbal score averages for SAT takers in six cohorts (1947, 1955, 1962, 1967, 1972, and 1975) as a starting point, an analysis was made of the Reading and History textbooks and related materials that were in commonest use at the 1st, 6th, and 11th-grade levels when these various SAT cohorts were getting their preparatory education. The materials were compared on the basis of several recognized measures of "readability," "difficulty," and "challenge."

Although this study proved inconclusive with respect to establishing direct correlations between the changing nature of the teaching materials and subsequent SAT scores, several of its broader findings are illuminating. By the measures adopted in the study, current 11th-grade texts are generally at what is considered a 9th-to-10th-grade level. The significance of this in relationship to the SAT experience is highlighted by the further finding that the reading passages and questions on the Verbal part of the SAT are – by these same measures – on an 11th-to-12th-grade and in some cases at a 13th-to-14-grade level of readability.

The study develops in statistical form the fact that a constantly increasing percentage of textbook space is taken up by pictures, larger print, wider margins, shorter words and sentences and paragraphs; the amount of exposition is decreasing, the amount of narrative going up. It also reports statistically on the extent (as indicated in the textbooks and related teacher materials) to which student testing is being reduced to an "objective answer" basis, meaning that students will find less reason to learn to write: "generally, the assignments in the Reading, History, and Literature textbooks [ask] only for underlining, circling and filling in of single words." Samples of good writing are accompanied in one text by the editor's (or publisher's) reassuring note to the students that while the selections "should be read and appreciated, they [are] not to be considered examples of writing that students should expect to attain themselves."

The study confirms what we know from the reports of textbook writers enjoined by publishers to "make it simple" and from the echoing reactions of better students that what they are reading at school is "simpler stuff than we read in the newspapers."

Recognizing that textbooks are not in themselves measures of good or bad teaching, or of how hard or easy a course is made, or of how much writing will be required, we nevertheless have a dual reaction to the reported changes in the nature of teaching materials. If textbooks are being written down simply in response to somebody's persuasion that students at various grade levels don't have what they used to have, this is greasing the toboggan – and is *itself* a cause of the SAT score decline. If, on the other hand, these changes reflect – as we suspect they do – an attempt to adapt education to the needs of *all* students, they represent an effort that reflects an authentic purpose but yields negative side effects by shortchanging some students.

The American education system is unique in its variety and its capacity to be useful to an extremely broad constituency, in which those who are going to take the SAT are a minority. We do not read the SAT score decline as an instruction that education in this country must or should be more rigid, more selective, more rejective, more uniform. Instead, the instruction is that education, especially secondary education, must become still more diversified, more varied – but without being watered down.

So in general we find that there *has* been a lowering of educational standards and that this *is* a factor in the decline in SAT scores. We conclude at the same time that the correction of the various elements in this situation requires the collaboration of teachers, students, parents, and the broader community in the establishment of standards that can be truly considered higher only as they recognize youths' essential diversity.

Staff and Facilities

The composition of the panel makes it poor critic or judge of the extent to which the decline in SAT scores may reflect simply teacher incompetence and administrative ineffectiveness, or, on the other hand, inadequate provision of educational facilities.

The sudden influx of students into the elementary and secondary schools during the 1960s and early 1970s resulted in a sharp increase in the demand for teachers and a considerable reliance on substitute teachers. The average years-of-experience figure for elementary school teachers was 13.3 in 1961; it dropped by 1966 to 10 years; then moved down in 1971 to about 8 years.

On the other hand, teachers' educational levels have been rising for the past 30 years.

Reported pupil–teacher ratios have been going down gradually but steadily in both elementary and secondary schools during the entire period relevant to the score decline: from 29.6 pupils per teacher at the elementary level in 1956 (for the country as a whole) to 23 per teacher in 1974; and from 21.2 to 19 students per teacher in the secondary schools. These figures are misleading in their lumping together of full-time classroom teachers and auxiliary personnel; the nationwide averages conceal the fact that in a good many school systems few teachers ever meet classes with fewer than 30 students in them. Although we do recognize that the compositional changes that have occurred place added burdens on teachers and argue for reconsideration of their responsibilities in order to achieve better writing and other desirable ends, it would be a mistake to attribute the SAT score decline to teacher overload.

Though there are no reliable objective measures of administrative competence, elementary and secondary school principals and superintendents have become increasingly men and women with both teaching experience and special training as administrators. Many of them now are political executives in the better sense of the term.

Teaching and administrative salaries have been increasing in comparative terms throughout this period and the general financial support base for education has been substantially expanded. It is clearly arguable that only part of the price tag on *real* equality of educational opportunity has yet been paid. We suspect that the inequitable distribution of funds for schools may be related to the problems of standards and quality, but we cannot claim to have investigated the matter. Yet to attribute the SAT score decline to inadequate financial support of education would both abuse the facts and worsen the ironical truth that this decline has already prejudiced the probably legitimate case for increasing this support.

These statistical facts—regarding teachers' years of experience and training, student–teacher ratios, salary levels, and the like—obviously involve, however, only superficial aspects of the issue that is now becoming the center of national debate regarding what to do about the quality of education. Despite increasing recognition of the role of the community and the family in learning's processes, attention is being centered on the qualification of teachers.

Over half the state legislatures have either enacted or are considering bills to condition the hiring of elementary and secondary teachers on their achieving statutorily prescribed grades on standardized teacher examinations. The legislative debates invariably invoke the decline of student scores on the SAT and other comparable examinations as support for these bills; the decline is introduced as evidence in litigation regarding the legality of this legislation.

The panel recognizes and in no sense minimizes the role of teachers and school administrators in what has happened here. Our best judgment is that their responsibility centers in their having made more concessions because of changing circumstances and demands—by tolerating excessive absenteeism, for example, and by themselves credentialing incompetence, by adopting less-demanding textbooks, by condoning little reading and less writing—than has been good for anybody involved. But this becomes a hard question of how much choice they have had, and of how the demands of a changing student clientele are best met.

We also recognize that there are compositional changes taking place in the teaching profession, involving the elimination of racial discrimination, that parallel in some respects the compositional changes among high school juniors and seniors and students going on to college. We see as much danger in compromising standards in the one case as in the other. Our increased understanding of the workings of standardized student examinations in such situations does not, however, commend exclusive reliance on the use of comparable examinations for teachers. Whatever may be wrong with teachers today is too important to be left to standardized examinations alone.

The Broader Learning Context

Most of us, doing poorly on a test, go through two stages in trying to figure out why it happened. With the SAT score decline there are four stages. Suspicious first of the test, next of the schools, then of the younger generation, we turn finally to look in the mirror. But the reflection is cloudy, our vision of ourselves only partly clear.

Most people accept the propositions that the population wave that engulfed the schools was of national making, that school absenteeism begins at home and corridor crime in the streets, that ill-considered promotions from one grade to the next trace at least as much to parental and community pressure or indifference as to principals' taking the easy way out. Yet there is virtually no statistical evidence of any *causal* connections between societal developments and SAT scores. There are delicate matters involved here—such as changes in the American family, the advent of television, and an embarrassing decade in American history.

Parents as Teachers

James S. Coleman, Christopher Jencks, and others have now established the relationship of various "family background variables" to educational outcomes, and we have taken account in Part Three of the unquestionable influence of some of these variables—socioeconomic status, for example—on SAT score averages during the period of changing composition of the test-taking population. But there is more here than that.

The period preceding and including the decline of the SAT scores has also been a period of marked change in the family.

Only one aspect of this change, involving family size, has been analyzed in terms of its possible relationship to college entrance examination scores. Several studies indicate that first-born children average higher scores on these tests than their siblings do and that the averages decline with each succeeding birth. This has been correlated with the decreasing score averages during the period when the post-World War II population wave affected this situation. Although the effect of multiple variables may be relevant, we find these research findings credible as far as they go.

More penetrating investigation of other possible causal factors in this situation has been handicapped by social sensitivity; the Student Descriptive Questionnaire asks

SAT takers how many siblings there are at home, but not how many parents. So research is only now beginning into the questions of what effect there may be on academic test scores as a consequence of broken homes, of there being one parent (or none) present instead of two, of the mother's working outside the home either by necessity or by choice, or of the father's allocation (again by either necessity or choice) of his time and interests.

We do know that in 1960 some 89 percent of all children under 18 in this country were living with both parents, and that this figure has dropped now to 80 percent. The number of children in other situations—living with one parent or none—is increasing at the rate of over 300,000 each year.

We know that the number of children from divorced families has doubled in the past 10 years, and that desertion rates are increasing.

We know that one obvious implication of women's exercising the now recognized right to equal employment opportunity is that they will be spending more time outside the home. The figures, which may be enemy to the truth because they tell only part of it, are that more than half of all women with children of school age (and 40 percent of those with children under 6) are now holding either full-time or part-time outside jobs; and these figures are rising rapidly.

What we *do not* know is what the relationship of these developments to children's learning may be—either under present circumstance or when appropriate adjustments are made to accommodate to a changing set of family life styles. The question usually left out is what the learning implications of alternative situations would have been in the particular case involved. The tendency is to look too much at the "status variables" in the picture, because they can be measured, and too little at the "learning process variables" that actually determine what kind of education takes place in any home as well as in any classroom.

In the panel's view, nevertheless, the matter of home learning is of critical importance in any attempt to identify the causes of the decline in academic test score averages. Although the key factors have not been identified—or even whether they involve the nature and amount of communication and of reading that goes on in the home, or the balance between concentration and distraction, or the mysteries of motivation and of self-image—there is probably more than coincidence between the decline in the SAT scores and the drop in the number of children living in two-parent homes.

Perhaps there is another related factor here. If there have been negative effects from changes in parents' functioning as teachers, they are probably paralleled by the effects of a weakening in traditional relationships between teachers at home and teachers at school. Most parents, regardless of the particular family situation, want to help in this teaching business as much as they can, but a lot of them feel they no longer know how. Changes in educational practice, as well as changes in family life styles, have probably contributed to a strain on the teacher–parent relationship, which may have had more effect than anything either partner has done alone. If this is true, as we suspect but cannot prove, it is an encouraging report from school book publishers that one of the largest new demands (next to the call for more basics-oriented and career-oriented materials) is for texts and teaching guides aimed directly at re-enlisting parent participation.

It is perhaps equally important to recognize that if little is known about the impact on achievement of changing family configurations and circumstance, still less is understood about the effects of the changing state of childhood and youth itself. There has been a steady lengthening of the period during which the family either cares for or seeks surrogate care for its younger members and exercises some form of jurisdiction over them. At the same time, there has been perceptible acceleration of young people's physical and sexual, and probably intellectual, maturation.

Are these opposing trends creating within youth, and between them and adults, tensions that manifest themselves in school in the forms of rebellious behavior and underachievement? The 17-year-old of 1977 is not the same in body or in mind as the 17-year-old of 1947 was. It can hardly be coincidence that problems of discipline and absenteeism appear at a time when changing life styles and values in adult society, earlier physical maturity, higher mobility, drugs, and the pill are all interacting.

None of this is translatable in the present state of the evidence into point declines on college entrance test score averages. Yet if the question is why those scores have been going down, few would respond without recognizing that part of the answer is almost certainly hidden in these gaps in present knowledge—about the effects of change on the whole meaning of family and youth.

Television—Test of the Modern World

By age 16 most children have spent between 10,000 and 15,000 hours watching television, more time than they have spent in school. When they reach the 1st grade, their average watching time is between 20 and 35 hours a week; this usually peaks at about age 12. The average time per child per day increased by about an hour between 1960 and 1970. Children are doing what their parents are; television now occupies about 40 percent of Americans' leisure time.

Is television a cause of the SAT score decline? Yes, we think it is. This cannot be proved, and we don't know *how much* a factor it is. By 1965, when scores started dropping, there were already television sets in 95 percent of all American homes; so ordinary research methods won't work here, for there are no non-television-watching control groups to use for comparison. Neither the difficulty of proof nor the impossibility of measurement, however, warrants diluting the answer. Television has become surrogate parent, substitute teacher.

What direct research there is on correlations between television watching and academic test scores is in fact entirely inconclusive. One earlier study comparing a television-watching Canadian town with another that was still at the radio-listening stage found T.V. Town's children had advanced vocabularies in the 1st grade, but that this advantage had disappeared by the 6th grade; there were no observable differences in academic achievement in grades 6 through 10. A 1954 series of studies of children in grades 3 through 6 in Evanston, Illinois, found that children in the lowest quarter of academic achievement averaged six hours more television viewing per week than those in the top quarter; but these studies took no account of other variables. In a sophisticated study conducted in Tokyo in 1970, an apparently significant correlation emerged at first between academic achievement scores and the test takers' reports of the amount of television they had been watching (the more television the lower the scores); but when controls were introduced for such variables as intelligence, creativity, and adaptability the separate significance of television watching disappeared.

So the panel's conclusions here must be plainly identified as essentially subjective.

We suspect that the three or four widely known and generally respected educational television programs designed for younger children are giving them a faster start with their letters and numbers. This may be at least one reason standardized tests show increases rather than declines in recent years at the first several grade levels.

So far as the scores on college entrance examinations are concerned, however, we are impressed and persuaded by two interrelated considerations. One of these involves the assumption that spending 10,000 to 15,000 hours on television's fare means a significant reduction in time, some of which would otherwise have gone into the development of the skills and aptitudes measured on college entrance examinations. An unquestionably considerable amount of time at the set used to go into homework and into reading and writing. To call television a thief of time is in a sense to beg the issue. Yet if developing the capacities that are measured by the SAT is taken as a value, the only question is whether television's larceny is petty or grand.

Another dimension of this first consideration involves the hypothesis that part of television's impact on education is that it raises children's expectation levels so that they are then discouraged by the comparative blandness of what they find when they open a book or encounter the next morning in the classroom. A member of the panel who is a teacher makes the point: "Sometimes I feel I'm competing with television stars who can sing and dance while they add and subtract and do the alphabet. I can't!"

The second set of considerations seem to us perhaps more important for what they suggest about the possible reach of television's implications than for anything specific. We have considered the apparently supportable working hypothesis that reading a line of script or type (as on a college entrance examination or in a textbook) involves a "linear, verbal, logical" function, which is performed in the left hemisphere of the brain, while watching something such as a television screen involves a "simultaneous, visual, affective" function – performed in the brain's right hemisphere.

Mindful of the dangers of a little learning, we simply note some of the questions those with larger but still very incomplete knowledge are also asking – without yet knowing the answers. Could this much (10,000 to 15,000 hours) functioning of one area of the brain alter the neural mechanisms of the mind – with *possible* effect on the handling of verbal materials? When so much more of children's time now goes into developing what they need for processing information in the visual mode (as in television) instead of the reading mode, would this result in lowered performance on the SAT test of verbal aptitude?

Could there be a suggestion here of a reason for larger declines on the Verbal part of the SAT than on the Mathematical part – in the possibility that geometry, for example, involves more a visual than a linear function, that mathematical problem solution is more simultaneous and holistic than linear, that the sense of algebraic equations comes from a grasp of the whole equation rather than from a linear view of one symbol following another?

Is part of the reason children's academic test scores are staying at previous levels or are increasing in the lower grades, but not in the higher, that there are closer similarities between television and the teaching methods used in the first few grades?

We know only that we stand on a frontier of knowledge that probably leads to partial answers to the questions we are considering, and that television's effects may in fact be so much broader than this one suggestion of them as to mock its unimportance.

The prospect is one of competing promise and threat. Traditional education and television are currently out of kilter, and the stakes in correcting this quickly seem to us higher than is generally realized. Yet if television's abuses have made it a "vast wasteland," its potential is of becoming learning's most fertile grove. We see this only as a question of when the family and the community will decide, as they have in the case of the schools, what kind of education they want this master but unprincipled teacher to give their younger members.

"I believe television is going to be the test of the modern world and that in this new opportunity to see beyond the range of our vision, we shall discover either a new and unbearable disturbance of the general peace or a saving radiance in the sky." This was E. B. White; the date was 1938. On that test, too, the scores have been declining.

A Decade of Distraction

When democracy thrives, Bernard Murchland points out, so does education, and when democracy is in crisis, so is education. Whatever application this truism may have to the SAT score decline is best suggested quickly and let go at that, for the facts are as obvious as the proof of any causal relationship is impossible.

The SAT scores dropped more sharply between 1972 and 1975 than during any other period. Most of the score decline on the ACT also came after 1970. It is hard to understand the suddenness and concentration of these changes—except for the realization that the students entering college during that period had gone through five or six years of national disillusionment, especially for young people, virtually unparalleled in American history.

It is arguable that the military draft, with its exemption of students, had a direct effect on the SAT score average along the course of its decline. Although our attempted analysis of this particular possibility leaves it in question, we suspect that there was a related but broader set of influences here. There is simply no way of knowing how much the trauma, between 1967 and 1975, of coincident divisive war (which youth had to fight), political assassination (of their particular heroes), burning cities, and the corruption of national leadership affected the motivations of the young people of that period—and whether there was consequent effect on their college entrance examination scores.

That concatenation of sad events unquestionably undermined respect for established institutions and processes, and this was manifested most overtly by young people. Because they were closest to education's institutions and processes, these were the focus of their protest. It was a time of extraordinary distraction, when it would have been hard for students to put the best that was in them into getting high marks on a college entrance examination.

This probably made quite a difference.

Motivation

The further the panel has gone into various aspects of this general question of reasons for the score decline, the more the matter of "motivation" has come up in varying forms relating not only to the taking of the college entrance examinations themselves but to virtually all aspects of the learning process.

A study commissioned by the panel, summarizing research in the area of achievement motivation, warns of the complexity of the subject and the difficulties of applying what is known about it to the limited information available to us. Yet we sense the motivational implications of a good many of the points that have been considered in other contexts.

It seems plausible speculation that as opportunities for getting into college have widened there may have been less concentration of student efforts on preparing for college entrance examinations. This would be in no way inconsistent with recognition of the fact that reducing anxiety improves some test takers' performance. The point is rather that there would appear to have been a lessening, for better or for worse, in the competitiveness or higher purpose that can be an incentive to excellence.

We have suggested earlier our persuasion that women's lower average scores on the Mathematical part of the SAT reflect a traditional stereotyping of roles and occupations, which has resulted in at least some young women's seeing no earthly point in developing that particular proficiency.

The differences in academic performance among students with different socioeconomic backgrounds have been widely analyzed in terms that put less emphasis on cognitive development than on training in independence, resourcefulness, confidence, willingness to take risks, and realism in setting aspirations levels.

Distinctions are drawn in some of the studies between the effect of various patterns of parental relationship (1) on what children learn, and (2) on the degree to which they are encouraged to learn.

Are ill-considered promotions, increased absenteeism, and "easier" textbooks causes or consequences of reduced student motivation? Both, we suspect.

To have noted some of the motivational elements that seem apparent in the effects of television is only to have suggested others of perhaps larger significance.

We have made the point, involving what is essentially motivation, that the extraordinary national events of the past 10 years may very well have had the effect of distracting young people from the pursuit of intellectual excellence.

It has been suggested that youth's changing job prospects and attitudes toward work may be important elements in the chemistry of their motivation to prepare for college entrance examinations. Although this is probably true as far as individual students are concerned, some fairly careful analysis of the timing of "labor market" changes over the past 10 years and changes in the test score averages discloses no significant relationship so far as any *net* effects are concerned. There have probably been different changes in this respect between female and male attitudinal patterns, and between those of higher- and lower-scoring groups of students. It is difficult to generalize about what effect varying degrees of encouragement and concern about job prospects may have in stimulating efforts to qualify for whatever may be available.

We note the possible motivational implications of the reports regarding two sets of questions on the Student Descriptive Questionnaire. One asks whether the test takers "want to receive help outside regular course work from the college you plan to attend. . ." going on then to list specific areas of possible assistance, including reading, mathematics, and writing. The answer is a resounding No in each of these three cases, by margins of about 8 to 1; and the percentage answering Yes has fallen off during the three years this question has been asked in its present form. (The only question showing any different reaction is one about help in finding work; a third of the test takers say Yes to this.) The panel members interpret this no-help-wanted reaction variously as a reflection of motivation, conditioned response, or a declaration of students' independence.

A good deal more significance may attach to the fact that there have been smaller than average declines during the past several years in the SAT scores of students who report on the SDQ that they have participated in various high school extracurricular activities (athletics; social clubs or community organizations; religious activities; music; journalism, drama, debating; student government; preprofessional or departmental clubs; or ethnic activities), and that they expect to do more of this in college. There are different views about the motivational elements involved in such participation, but the item has perhaps special interest in view of the suggestion that engaging in this kind of activity is a particularly reliable index to young people's longer-run future accomplishments. We commend further inquiry by the College Board and ETS into this area. It could possibly have considerable significance in suggesting not only basic elements in the reasons for the SAT score decline but also alternative forms of predictive testing.

We have noted in an earlier section of the report (Part Three) the possibility that there are motivational elements reflected in the fact that students taking Achievement Tests in conjunction with the SAT show a significantly less-than-average score decline pattern on the SAT itself (at least on the Mathematical part).

These would be, even at best, only possible footnotes to a motivational analysis of the SAT score decline, which we are not competent to make but which we commend as warranting further consideration. Such an analysis would not be limited to students' attitudes about tests but would include consideration of the motivational elements in the interrelationship between formal education and its broader societal context. It appears to us to be not just youth's motivations that are involved but also those of the society.

With respect to not only this motivational point but the preceding points as well, the panel notes that the frustrating limitations on understanding of the elements contributing to the "pervasive" influences on SAT-measured competence will be removed only when two additional research steps are taken: first, the extension of the Student Descriptive Questionnaire's scope to cover individual test takers' circumstances outside their formal education; second, the undertaking of longitudinal studies permitting the following up of sample cohorts of young people from childhood to maturity. The society has drawn back from this on grounds of concern about rights of privacy. But these can be fully protected. The larger problem is to overcome sensitivity about inquiring into areas in which the answers may be embarrassing because of what they indicate about such influences as poverty, prejudice, and the functioning of various institutions, including particularly the schools and the family.

Also, it is important to note that a limited instrument such as the SAT should not become the sole thermometer for measuring the health of schools, family, and student. It tells us nothing about young people's honesty and integrity, about whether they care about each other, or about a lot of other things that matter more than test scores. It would be a mistake to think about such major institutions as the schools and the family solely in terms of their effects on test scores.

In the View of Others

The panel has benefited from a generous volunteering of views regarding reasons for the SAT score decline. Although most of these comments are incorporated or at least reflected in various other sections of this report, this is not true of all of them. None of the suggestions received has been disregarded, and it is appropriate to note here those brought to our attention that we have either not pursued as fully as may be warranted or have followed up to the point of concluding that they are not significant.

Many of these reactions relate to the schools and suggest at least partial attribution of the decline to:

- "New math."
- Reading programs with insufficient emphasis on phonetics, vowels, consonants, and so forth.
- The expenditure of funds on school facilities instead of on teachers.
- Rising costs of postsecondary education, discouraging some superior students from applying to the characteristically more expensive colleges and universities, which require SAT scores.
- The independent school board system, which makes it impossible to set and require national achievement standards.
- Increasing legislative involvement in education, "tying the hands of school personnel with respect to discipline."
- Excessive reliance on clerical aids.
- The influence of "the soft pedagogical left," which "believes expressiveness is an adequate substitute for thinking and knowing, and which views leniency as a kindness to the underprivileged."
- The increased numbers of married female teachers.
- The "influx of male teachers and principals in the 1960s into a traditionally female domain."

Because of the frequency of suggestions that the score decline is attributable to "the introduction of experimental teaching methods," (or to "open classrooms in which students choose activities in nongraded groups and work largely without teachers," etc.) the panel made particular efforts to explore this general area. Two different investigations proved generally inconclusive. Their indicated results are confirmed, however, by a federally sponsored study of some 30,000 students at the primary and elementary levels in 13 school districts in nine states, recently (December 1976) completed by the American Institutes for Research. The AIR study was based on a three-year follow-up of students who took the Comprehensive Test of Basic Skills in 1970-71 at the 1st, 4th, or 6th-grade levels. "The single most important and well-documented finding," the AIR report concludes, "was the lack of either substantial or consistent association between student achievement and overall level of innovation across the grades." We find no evidence of any causal relationship between what are commonly referred to as "experimental teaching methods" and the SAT score decline.

The panel also looked into the suggestion that the introduction of more experiential (outside-of-school) training in the high schools may have contributed to a lessening development of students' abilities in the basic skill areas covered by the SAT. What evidence there is points in the other direction. One of the SDQ questions is about the part-time work the test takers have been doing. It has been going up slightly since 1973. Yet test takers reporting 1 to 15 hours of outside work a week average higher SAT scores than those doing no such work at all, and their score *declines* over the past four years have been less than the overall average. (Students reporting more than 15 hours of outside work show generally lower-than-average SAT scores, particularly Verbal scores. We assume, though without knowing, that this is the lower-socio-economic-status group.)

A related question involves whatever shift may be taking place at the secondary school level away from liberal arts and toward more vocational and technically oriented training. Here again there is only totally inconclusive evidence about any possible effect on the SAT score averages. The SDQ reports confirm the fact that test takers enrolled in "career" courses of study in high school (which is a very small percentage, about 6 percent) average much lower scores on the SAT than do those enrolled in "academic" courses. The differences are about 100 points on both the Verbal and the Mathematical parts of the test. There has been only a negligible shift since 1973, however, in the percentage of SAT takers in the two groups; and the score declines have been almost identical.

Although the panel's attention has been directed repeatedly to the facts of increased schoolground violence and crime and juvenile alcoholism and drug addiction, we can add nothing here to what common knowledge and common sense already establish. These aberrations obviously affect not only the individuals directly involved but the broader educational process, and they have been increasing as the SAT scores have been going down. It seems worth noting only that these matters are consistently raised in terms of what is happening in the schools and among children. The problem is obviously broader than that. So far as the issue is why the SAT scores have been going down, the question is: What is causing the delinquent behavior?

With respect to the "forced busing" question, also frequently raised, we conclude from what information it has been possible to obtain that in those relatively few instances in which busing has meant large-scale disruptions of one kind or another there has been an immediate negative impact on the educational process, although this has been characteristically of short duration. There is no way of determining or quantifying its *net* effect. Its incidence has in any event been restricted to so small a proportion of the national student population that it could have had at most only very slight effect on the SAT score averages.

There have been a number of suggestions that teacher organization and the development of collective bargaining have affected adversely both the amount of time teachers spend with students and their attitudes toward the profession. We have looked into this only to the extent of determining that there is no evidence available to permit an objective judgment one way or the other. This has seemed a sufficient basis for not going further into an issue on which the panel members' views would probably be about as divided as are those of the broader public.

So far as suggestions regarding possible causal factors outside the school system are concerned, the panel has taken particular notice of a number of suggestions that the score decline is attributable to health factors of one kind or another: to the increased use of chemical additives in foods, for example; to an increased resort to induced labor in pregnancy (with resultant brain damage); or to the effects of the uses of various forms of anesthesia in connection with childbirth. A study was commissioned to determine, broadly, the possibilities of a medical basis for the declining SAT scores. Its advice to the panel is, in short, that nothing in the available evidence warrants identifying these factors with the score decline, and that the evidence of stable or even increasing score averages in tests administered at the primary grade levels denies the suggestion that the test results at the higher grade levels can be properly traced to medical or health factors that have their largest impact during earlier childhood years.

Other suggestions tie the test score decline to increased pregnancy among younger women and rising abortion rates.

Still others attribute the decline to "a growing rejection of traditional Western religions" and to a concomitant turning to "religions of the East, drug-related religions, mysticism, witchcraft, astrology," or, with different emphasis, to a growing "preference for fantasy over reality" and "celebration of the ideology of irrationalism, in which knowledge is attained through intuition, inspiration, and revelation."

Concern about a "crisis in values" is widely expressed, but with varying emphases: on "subversive political activity by Communists"; on the promotion of pornography; on the "grabbing of political power by the NAACP, NEA, CLU and others . . . [so that] public school children are misguided through negative education, . . . groomed in total disgust with the nation, family, and self"; on the "impact of the counterculture with its hostility to reason, to science, to technology, to industry, to the work ethic"; on a "revolution in values, including a decline in the Protestant ethic . . . , the idea of success, . . . the idea of work itself"; a "tidal wave of moral bewilderment, concern, and resentment."

Although some of these suggestions seem to the panel to offer little in themselves as far as explaining the decline in the test scores is concerned, we imply nothing one way or the other here about the broader values they reflect. In general we find the sum of these contributions substantially helpful in suggesting the character of a period, covered by the score decline, which has been an unusually hard one to grow up in.

Part Five. Summing Up

If you turned to this concluding section for a quick and easy understanding of the panel's views on the decline of test scores, you are indulging in a practice like some of the educational shortcuts that may have contributed to the decline. This is a complex subject, and our views of it are filled with nuances, qualifications, and some doubts. The quality of our judgments is conditioned by the quality of the evidence. Without a reading of the report, the summary may suggest a simplicity that is unfair to an important subject.

Looking first at the technical aspects of the SAT score decline, the panel has determined that the decline does not result from the test having become, in terms of common usage, "harder." More precisely, our analysis of the equating and scaling procedures that have been followed here indicates that the decline has been several points larger than is shown. To the extent, nevertheless, that the SAT is held out as providing an essentially constant measure of levels of aptitude in the areas it covers, permitting reliable comparisons over periods of time, we find it to have been maintained with superior competence and fidelity.

We find that, despite the decline in score averages, the SAT continues to be a valid instrument for helping to determine how well those taking it will do in their college courses, especially those in the first year. Its predictive validity has in fact increased slightly over the past several years.

We commend to the College Board and ETS, nevertheless, the undertaking of a broader inquiry into the function of tests at the passage point between high school and college. Such an inquiry would not be limited to what conventional testing procedures measure regarding traditionally emphasized competencies; nor should it be confined to an evaluation based on examinations and statistics comparing the effectiveness of learning today with what it was in the past—by standards the past adopted. That inquiry would proceed from a determination of how much this country wants to develop of the limitless human resource, which is what education is all about.

The panel members share strongly what has now become national concern about the implications of the declining test scores, about the increasing (or at least increasingly recognized) signs of functional illiteracy, and about an apparent deterioration in the use of the English language. Yet we share, too, the sobering realization that a good many of the younger members of our families seem to know quite a lot more about quite a few things than we did when we were their age. Even as we pursue this inquiry into what is apparently going wrong as reflected in the decline of the SAT scores, we urge that a broader look be taken at the whole picture—including whether we are testing in the best possible way whatever ought to be tested.

Turning then to the broader aspects and implications of the SAT score decline, our assessment of this continuous 14-year drop in averages is that it is unquestionably significant. Particularly when the SAT record is set beside the broader pattern of comparable declines on other standardized academic tests, it emerges as a development warranting careful attention by educators and by everybody interested in education.

It is in no way inconsistent with this appraisal to point out that these declines cover a period of notable extension and expansion of educational opportunity in the United States, coincident with the experiencing of virtually unprecedented turbulence in the nation's affairs. The record may suggest as much about youth's inherent resiliency and the resourcefulness of the formal educational process under unusual circumstances as about deterioration in either personal or institutional fibers. We have wondered sometimes in the course of our inquiry why the score declines haven't been larger.

In general, the causal factors apparently involved here fall into two categories so different that it seems helpful to think in terms of what are virtually two score declines. One reflects primarily changes in the SAT-taking population; these score averages measure a different and broader cross section of American youth from the group they measured 20 or 15 or even 10 years ago. The reasons for the other aspect of the decline are more elusive; they include the apparently pervasive influences, affecting virtually all groups of students alike, of changes in the practices of the schools and in the American social fabric.

Most—probably two-thirds to three-fourths—of the SAT score decline between 1963 and about 1970 was related to the "compositional" changes in the group of students taking this college entrance examination.

That was a period of major expansion in the number and proportion of students completing high school, resulting only in part from the post-World War II population wave, which came along then. The rest of the growth reflected the deliberate national undertaking during that period to expand and extend educational opportunity—by reducing the high school drop-out rate, by trying to eliminate previous discrimination based on ethnicity or sex or family financial circumstance, and by opening college doors much wider.

In the panel's view, it would be a deceptive misstatement to describe this first cause of the SAT score decline as being simply that increased percentages of lower-scoring groups of students began taking the test. The cause lies rather in whatever the reasons may be for this lower scoring. What the decline reflects is the incompleteness so far of the national undertaking to afford meaningful equality of educational opportunity. This leaves the question of whether a 75 percent cross section of all young people *can* be brought up to the 11th- or 12th-grade academic attainment level previously achieved by 50 percent of them. Part of democracy's sustaining notion is that they can be.

It is important to recognize in a related connection that to the extent the lower SAT averages reflect an expansion of the test-taking population to include a different mix of young people, the decline indicates nothing about changing abilities or aptitudes of this age group *as a whole*. We simply don't know what the *net* effects here have been.

Even during this first period, however, there were emerging signs of more "pervasive" influences or forces, going beyond any "compositional" changes, which were having an effect not only on the overall SAT averages but also within the various groupings of test takers.

From about 1970 on, the composition of the SAT-taking population has become comparatively more stabilized with respect to its economic, ethnic, and social background. Yet the score decline continued and then accelerated; there were particularly sharp drops during the three-year period from 1972 to 1975. Only about a quarter of the decline since 1970 can be attributed to continuing change in the make-up of the test-taking group. With a handful of exceptions, the drop in scores in recent years has been virtually across the board, affecting high-scoring and lower-scoring groups alike.

This second set of factors contributing to the SAT score decline can be summarized only in broad terms and with full recognition of two related qualifications. First, any attempt to isolate developments in the schools from those in the society at large turns out to reflect principally the inclination to institutionalize blame for whatever is going wrong; the formal part of the learning process cannot be separated from its societal context. Second, to the extent these causal factors are understood at present, they are inextricably interwoven with each other: any pointing to one development or another as if it were *the*, or even *a*, cause of the decline is invariably misleading.

As already noted, we think that two-thirds to three quarters of the score decline from 1963 to 1970 and about a quarter of the decline since 1970 were caused by complex interacting factors relating to the changing membership in the population tested. Overall this suggests that about half of the decline is properly traced to these factors. The remainder seems to us identifiable in large part with six other sets of developments:

- *One.* There has been a significant dispersal of learning activities and emphasis in the schools, reflected particularly in the adding of many elective courses and a reduction of the number of courses that all students alike are required to take. This has been true particularly in the English and verbal skills area.

In the panel's judgment, any broadside condemnation of "more electives" is mistaken. Many of these courses are designed to interest and motivate students, and they are not properly considered as having by their nature a negative effect on basic learning. We would not recommend any single formulation of subject matter or teaching method, for it is clear that both traditional and innovative approaches to learning can produce good results—or bad. We are inclined to believe, nevertheless, that probably well-intentioned change has reduced the continuity of study in major fields with consequent effect on the development of verbal and (to a lesser extent) quantitative relations skills, and that too large a proportion of the curriculum changes in recent years has been accompanied by a tendency to avoid precise thinking and the demands it makes on both students and teachers. The SAT score decline probably reflects in part the effects of the schools' placing reduced emphasis on steady growth in verbal and mathematical competence.

We attach central importance to restoring the traditions of critical reading and careful writing.

- *Two.* There is clearly observable evidence of diminished seriousness of purpose and attention to mastery of skills and knowledge in the learning process as it proceeds in the schools, the home, and the society generally. This takes a variety of apparently disparate but actually interrelated forms: automatic grade-to-grade promotions, grade inflation, the tolerance of increased absenteeism, the lowering of the demand levels of textbooks and other teaching and learning materials, the reduction of homework, the lowering of college entrance standards, and the inclusion of "remedial" courses in postsecondary education.

Each of these issues presents its own quandary. We are *not* suggesting simplistic "solutions" through which all students are treated alike by being held in a grade until they reach a common standard, suspended from schools as a penalty for absenteeism, subjected to the same more demanding reading materials, overloaded with homework, confronted with some national common denominator of college entrance, or denied needed help in skills development in college.

Each of these problems has developed in response to the wider spectrum of interests and abilities the schools and colleges are now trying to serve. In a sense the schools may have tried so hard to accommodate the special needs of new and unfamiliar students that these very students along with others have been ill served by not being held to demanding expectations of performance. The lowering of teaching sights is the wrong answer to whatever may have been the consequences of the expansion and extension of educational opportunity. The only right answer is to vary the instructional process still more to take account of increased individual differences, but without lowering standards—which we recognize as a form of magic, but one that has been performed in this country for a long time.

- *Three.* Particularly because of the impact of television, but as a consequence of other developments as well, a good deal more of most children's learning now develops through viewing and listening than through traditional modes. Little is known yet about the effects of this change, including its relationship to performance levels on standardized examinations.

We surmise that the extensive time consumed by television detracts from homework, competes with schooling more generally, and has contributed to the decline in SAT score averages. Yet we are convinced that television and related forms of communication give the future of learning its largest promise. The most constructive approach seems to us to depend less on limiting the uses of these processes than it does on the willingness of the community and the family to exercise the same responsibility for what is taught and learned this way as they have exercised with respect to older forms of education.

- *Four.* There have unquestionably been changes, during the period relating to the score decline, in the role of the family in the educational process. Social sensitivity has precluded thorough inquiry into this area, so that only the readily observable structural changes can be noted: the rapidly increasing number and percentage of children, for example, in less than complete families. While evidence is not available to determine the effect of these changes on students' college entrance examination scores, our conjecture is that it is negative.

- *Five.* The concentration of the score declines in the three-year period between

1972 and 1975 leads the panel to suspect strongly that one important element here was the disruption in the life of the country during the time when those groups of test takers were getting ready for their college entrance examinations.

- *Six.* For whatever combination of reasons, there has been an apparent marked diminution in young people's learning motivation, at least as it appears to be related, directly and indirectly, to their performance on college entrance examinations. Although this may be largely only another dimension of the preceding points, it is perhaps most significant of all that during the past 10 years the curve of the SAT scores has followed very closely the curve of the entire nation's spirits and self-esteem and sense of purpose.

So there is no *one* cause of the SAT score decline, at least as far as we can discern, and we suspect no single pattern of causes. Learning is too much a part of Life to have expected anything else.

It would be too bad, furthermore, if our concentration on the implications of a decline in the statistical averages on a set of standardized examinations should seem to ignore how incomplete a measure this is of either educational or broader human purpose. While we ask why the scores on college entrance examinations have gone down, T. S. Eliot's probing goes much deeper: "Where is the learning we have lost in information? Where is the understanding we have lost in knowledge? Where is the life we have lost in living?"

Yet in the panel's view of it all, the fact of the hard asking—of both kinds of questions—offers new promise of new answers. We find nothing in the record we have reviewed to discourage the conviction that learning in America can be made all that is hoped for it. What is clearest is the reflection, in the reactions to these test scores and to the poet's lament alike, of renewed purpose to implement these hopes. The future continues to seem a good idea.

Reference Notes

The panel has relied in large part on studies undertaken and reports prepared at its special request. These are being published as appendixes to this report. An annotated listing of them follows these Reference Notes, as does a more comprehensive Bibliography of all sources on which the panel has relied. The specially prepared papers and reports are identified in the Reference Notes by an asterisk. They are all the work, except where otherwise indicated, of the staff of Educational Testing Service or the College Board.

Part One. The Scholastic Aptitude Test and the Score Decline

The description of the SAT score pattern is based largely on *A Summary of SAT Score Statistics for College Board Candidates** (1976) prepared for the panel by Rex Jackson. Additional sources are indicated in the footnotes to Table 1.

Sample questions from the various sections of the SAT appear at the back of this report, following the Bibliography.

Part Two. An Unchanging Standard

As It Was in the Beginning:

The technical details of procedures followed in making scaled scores comparable over time and equating new editions of the SAT with previous editions are set out in *The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests,* published by the College Board in 1971. For technical descriptions and assessments made under independent auspices, see *The Seventh Mental Measurements Yearbook,* edited by Oscar Buros (1972).

The panel's conclusions regarding the effectiveness of the scaling and equating procedures are based on two studies, one by Christopher Modu and June Stern, *The Stability of the SAT Score Scale** (1975), the other by Albert Beaton, Thomas Hilton, and William Schrader, *Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates between 1960 and 1972** (1977). The panel's review of possible "item obsolescence" in the test items which reappear in different forms and are used for equating was facilitated by the analysis by James Braswell and Nancy Petersen, *An Investigation of Item Obsolescence in the Scholastic Aptitude Test** (1976).

The discussion of the predictive validity of the SAT is based on the studies made by the colleges participating in the Admissions Testing Program Validity Study Service, the results of which have been synthesized by Susan Ford and Sandy Campos, *Summary of Validity Data from the Admissions Testing Program Validity Study Service** (1977).

The development and administration of the SAT are reported on in a series of four articles by Jim Landers and J. Stryker Meyer in *The Sunday Times-Advertiser* and *The Evening Times,* Trenton, N.J., March 28-31, 1976.

A Changing Context?

The panel's consideration of the "relevancy" issue was illuminated by letters received from Terry Herndon, executive director of the National Education Association, and Professor Robert Bannister of Swarthmore College.

The reference to the continuing correlation of SAT scores with high school grades is based on an analysis by Rex Jackson, *Correlations of SAT Scores with High School Record** (1977).

The student valedictorian story is reported by Stephen Klaidman in *The Washington Post,* August 15, 1976.

Cross sections of current adverse criticism of educational testing are contained in several recent symposiums: Association for Childhood Education International, *Testing and Evaluation: New Views* (1975); American Federation of Teachers, "To Test or Not To Test: An Examination of the Educational Testing Controversy," *American Educator,* Winter 1977; National Education Association, "Special Feature on Testing," *Today's Education,* March-April 1977; *The Myth of Measurability,* edited by Paul L. Houts (1977). See also Cynthia Parsons, "SAT + ACT = College Tests in a Mess," *The Christian Science Monitor,* March 29, 1976; B. Bruce Briggs, "The Great Classroom Debacle," *The Wall Street Journal,* July 20, 1976. The panel notes these criticisms without any suggestion of its views regarding them.

Part Three. The Two Score Declines

Compositional Change:

The data used at the beginning of the section to describe youth and student population changes are from Table 1 of the report. The discussion of changes in the proportions of students graduating from high school and going on to college is based on the historical series maintained by the Bureau of the Census, recording the percentage of 1,000 5th-graders who are retained in school at each grade beyond that, who graduate from high school, and who enroll in college (see *Digest of Educational Statistics,* 1975 edition, Table 10, p. 14, National Center for Educational Statistics). The statistics in Table 1 cannot be used to derive school retention rates since any one school class is composed of youth of varying ages, while the population figures in Table 1 cover only 18-year-olds.

The composite analysis of SAT, Project TALENT, and National Longitudinal Study data was made for the panel by Albert Beaton, Thomas Hilton, and William Schrader, *Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates between 1960 and 1972** (1977). Data from the American Council on Education's annual Freshman National Norms Study and from the ACT Student Profile reports were also used to examine compositional changes. See Richard L. Ferguson and E. James Maxey, *Trends in the Academic Performance of High School and College Students,* American College Testing Program Report No. 70, January 1976; E. James Maxey, *Trends in the Academic Abilities, Background Characteristics, and Educational and Vocational Plans of College-Bound Students,* American College Testing Program Report No. 74, May 1976.

Changes in college-going patterns as they affect SAT scores are analyzed in the study by William Schrader, *Distribution of SAT Scores to Colleges as an Indicator of Changes in the SAT Candidate Population** (1976), and in Samuel S. Peng's paper, "Some Trends in the Entry to Higher Education: A Comparison between NLS and Project TALENT" (1976).

The influence of changes in the extent of repeating the SAT is explored in Rex Jackson's study, *An Examination of Declining Numbers of High-Scoring SAT Candidates** (1977).

Regarding reliance on cram courses, see *Effects of Coaching on Scholastic Aptitude Test Scores* (College Entrance Examination Board, 1965); Lewis W. Pike and Franklin R. Evans, *Effects of Special Instruction for Three Kinds of Mathematics Aptitude Items* (College Entrance Examination Board, 1972); Bart Barnes, "Area Pupils Attend Special Cram Courses to Prep for SAT," *The Washington Post*, October 18, 1976; Linda K. Nathan, "'Improving' Law School Aptitude Is Big Business," *The New York Times*, November 14, 1976.

Pervasive Change:

Evidence of the relatively more stable makeup of the SAT population and of the pervasive nature of much of the decline during the period between 1973 and 1977 has been obtained from a computer printout of the Student Descriptive Questionnaire data. A generally corroborative picture of experience with respect to the ACT is provided by *The High School Profile Report, 1975-76*, issued by the American College Testing Program.

Further and more detailed information on the nature of the SAT score decline in this period is contained in the papers by Rex Jackson on *An Examination of Declining Numbers of High-Scoring SAT Candidates** (1977) and on *Trends in Mean SAT Scores for Students from the High School Classes of 1971-72 and 1975-76 Belonging to Selected School and College Groups** (1977); in Thomas Donlon and Gary Echternacht's report on *A Feasibility Study of the SAT Performance of High-Ability Students from 1960 to 1974 (Valedictorian Study)** (1976); in James Nelson's *A Review of Data Available regarding Family Income and Financial Aid Characteristics of Students** (1976); and in June Stern's analysis of *Selected Percentiles for Scholastic Aptitude Test Scores (1966-67 through 1975-76)** (1977). For a state-level study reaching similar conclusions about the pervasive nature of the score decline in this period see Janice J. Weinman, *Declining Test Scores: A State Study*, Massachusetts Department of Education (1977).

Through Other Looking Glasses:

The panel's consideration of the experiences of other major national or statewide testing programs is based on the standard published reports of the sponsoring agencies and on studies that include summaries and analyses of a variety of other tests. Particular note is made of the analysis of other test score experiences by Annagret Harnischfeger and David Wiley in *Achievement Test Score Decline: Do We Need to Worry?* (1975); the ACT research report by Leo Munday on *Declining Admissions Test Scores*, American College Testing Program Report No. 71, February 1976; a report by Frank E. Armbruster, *The U.S. Primary and Secondary Educational Process* (Hudson Institute, 1975); the analysis by John Flanagan and Steven Jung, *Progress in*

Education: A Sample Survey (American Institutes for Research, 1971); John Flanagan's report at the American Educational Research Association meeting in 1976 on "Changes in School Levels of Achievement: Project TALENT Ten and Fifteen Year Retests"; an exerpt* from the paper, *A Summary of Score Change*, delivered by T. Anne Cleary and Sam McCandless at the 1977 annual meeting of the Northwest Association of Schools and Colleges.

Data regarding PSAT/NMSQT scores are presented in Rex Jackson's study *Mean Scores for PSAT and PSAT/NMSQT Junior Candidates* (1976).

The discussion of Achievement Test candidates' scores on both those tests and the SAT is based on June Stern's *Table of SAT and Achievement Test Scores for Samples of Candidates Taking Achievement Tests 1966-67 to 1975-76** (1977).

Part Four. Circumstantial Evidence

In the Schools:

Courses of Study. As indicated in the report, extensive reliance has been placed by the panel on the studies by Harnischfeger and Wiley and by Weinman, noted above. See also the *Survey of Curricula Offerings*, prepared by the Massachusetts Department of Education, 1967-68, 1970-71, 1972-73, and 1975-76. The California picture is described in a series of three articles by Jack McCurdy and Don Speich in *The Los Angeles Times*, August 15-17, 1976; the panel has not made an independent check of the reported facts. Changes in courses taken have also been traced through special tabulations of the Student Descriptive Questionnaire data; these data are subject to the possible influences on student self-reporting.

The panel has not made an independent study of the extent to which there have been declines in reading and writing. Reference is appropriate, however, to the report prepared for the U. S. Office of Education by ETS and the University of Indiana (R. Farr, J. Turnman, and M. Rowls) on *Reading Achievement in the United States: Then and Now* (1975); Donald Graves, "Writing: An Endangered Species," *The Common*, March 1977; A. Bartlett Giamatti, "The Language Barrier: Why Students Can't Write," *Yale Alumni Magazine*, January 1976; "Why Johnny Can't Write," *Newsweek*, December 8, 1975. A study of the federally financed Follow Through program, entitled *Education as Experimentation*, has been made recently by Abt Associates for the Office of Education; not yet available for general distribution, its conclusions are suggested by Lawrence Feinberg in "Basic Teaching Methods More Effective, Study Says," *The Washington Post*, June 20, 1977.

Learning Standards. Average daily attendance data are derived from the *Digest of Educational Statistics* and *A Century of Public School Statistics* (1975), both prepared by the National Center for Educational Statistics. The National Association of Secondary-School Principals' report, on "Student Attendance and Absenteeism," is in *The Practitioner*, March 1975.

The panel's discussion of textbook content is based on a study made by Professor Jean Chall of Harvard University, *An Analysis of Textbooks in Relation to Declining SAT Scores** (1976). See also the *Report on a National Study of the Nature and the Quality of Instructional Materials Most Used by Teachers and Learners*, Educational Products Information Exchange Report No. 76 (1977).

Staff and Facilities. Information on teacher experience and education and on pupil-teacher ratios is based on data developed by various offices of the Department of Health, Education, and Welfare, particularly by the National Center for Educational Statistics (*The Condition of Education* [1975-77]); most of it is summarized in Harnischfeger and Wiley, noted above.

The Broader Learning Context:

Parents as Teachers. The specific references are to James S. Coleman's *Equality of Educational Opportunity* (1966) and to Christopher Jencks' *Inequality: A Reassessment of the Effect of Family and Schooling in America* (1972). A large number of studies of the comparative importance of family background and school influences have appeared subsequently (see Edward Kifer, "The Relationship between the Home and School in Influencing the Learning of Children," a paper delivered at the Pre-Convention Conference on Research, National Council of Teachers of English, Chicago, Illinois, November 23, 1976).

Regarding the effects of family size and sibling order on academic tests, the panel had available to it the studies of R. B. Zajonc, "Intellectual Environment and Intelligence" (unpublished paper, 1975) and "Birth Order and Intellectual Development" (*Psychological Review*, 1975) as well as Hunter Breland's *Family Configuration Effects and the Decline in College Admissions Test Scores: A Review of the Zajonc Hypothesis** (1976). See also Carol Tavris, "The End of the IQ Slump," *Psychology Today*, April 1976.

The Condition of Education (1975-77) (National Center for Educational Statistics) provides a historical series on children under 18 who are living with two parents, one, or none. See also Urie Bronfenbrenner, "The Calamitous Decline of the American Family," *The Washington Post*, January 2, 1977.

The report of trends in published teaching materials is from Barbara Thompson Howell's presentation to the New Jersey Association of School Administrators, Atlantic City, New Jersey, May 12, 1977.

Television—Test of the Modern World. We relied heavily on panel member Wilbur Schramm's report to us on *Television and the Test Scores** (1976).

The conjecturing about the functioning of the mind is based on suggestions received from Professor Bikkar S. Randhawa of the University of Iowa. There is help for understanding, too, in a symposium entitled "Inside the Brain: The Last Great Frontier," *The Saturday Review*, August 9, 1975; *Television and Children, Priorities for Research*, a report of a conference sponsored by The Ford Foundation at Reston, Virginia, November 5-7, 1975; "What TV Does to Kids," *Newsweek*, February 2, 1977. But see also Max Gunther, "How Television Helps Johnny Read," *TV Guide*, September 4, 1976.

The E. B. White quotation is from his column "One Man's Meat," *Harper's Magazine*, 1938.

A Decade of Distraction. At the panel's request, R. H. Glover prepared two contextual maps of *Major Societal Changes in U.S. (1933-44 and 1945-75): Contextual Mapping** (1976). The influences we refer to particularly are described in the chapters entitled "Up against the Wall" in William Manchester's *The Glory and the Dream* (Boston: Little, Brown and Company, 1974).

Motivation:
The paper prepared for the panel is by Professor David Winter of Wesleyan University, *Motivational Factors in the SAT Score Decline** (1976).

We have also relied to a limited extent on the Student Descriptive Questionnaire data, but with realization of the various elements that may affect the test takers' answers to these questions.

In the View of Others:
These views have been distilled from a large number of letters, personal conversations, and publications by Yvonne Wharton, in *List of Hypotheses Advanced to Explain the SAT Score Decline** (1975).

The two studies of the history of SAT scores in various types of schools are Gary Echternacht's *A Comparative Study of Secondary Schools with Different Score Patterns** (1976) and Rex Jackson's *Comparison of SAT Score Trends in Selected Schools Judged To Have Traditional or Experimental Orientations** (1976). The panel also drew on the study by the American Institutes for Research, *Impact of Educational Innovation on Student Performance* (1976).

Annotated List of Studies and Papers

The panel asked for a variety of studies and papers to assist in understanding the score decline. These papers proved useful for that purpose, even though the panel did not use all the material in them or necessarily agree with all the conclusions. They are published as a separate Appendix to the panel's report, so that this information will be available to others who wish to go more deeply into the subject. What follows is an annotated list of these commissioned studies and papers.

Arnold, C. B. *Could There Be a Medical Basis for the Declining SAT Scores?* January 1977, 8 pages.

A brief summary and analysis of the medical and epidemiological literature dealing with nutrition, drugs, genetic and prenatal conditions, labor and delivery, child development, health status, and health care over several decades in the United States. It concludes that the decline in SAT scores is probably not the result of disease processes or physical environmental factors.

Beaton, A. E.; Hilton, T. L.; and Schrader, W. B. *Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates between 1960 and 1972.* January 1977, 81 pages.

Two major social science research efforts—Project TALENT in 1960 and the National Longitudinal Study in 1972—obtained extensive data on ability and other significant characteristics for national probability samples of high school seniors. Both surveys made follow-up studies to determine which of these seniors had entered college in the year following high school graduation. This study was designed to use these two exceptional data bases to document changes between 1960 and 1972 for high school seniors, college entrants, and the SAT-taking population. The data of the national surveys were supplemented by the results of a special equating study and by a search of SAT files for about 20,000 members of the TALENT sample.

The study was concerned with three groups—high school seniors, college entrants, and SAT takers. Because reading scores were available for all three groups, it was possible (after equating the reading tests used) to measure changes in reading ability for all three groups. In addition, it was possible to study subgroups of the three main groups. The subgroups were defined on the basis of each of the following characteristics: age, sex, father's education, mother's education, father's occupation, mother's occupation, family configuration, high school curriculum, and expected college major field.

The main conclusions of the study are as follows.

1. All three groups showed a decline in reading ability between 1960 and 1972. The

decline for SAT takers, however, is markedly greater than that for the other two groups.

2. A much greater increase in the proportion of low-ability than of high-ability students who took the SAT appears to be the predominant source of the SAT score decline between 1960 and 1972. The decrease in ability level of high school seniors during this period also contributed to the decline.

3. There is some evidence that SAT scores earned in 1960 and 1972 are not precisely comparable. The data suggest that the actual decline in average verbal ability of SAT candidates from 1960 to 1972 was somewhat greater than the SAT scores indicate.

4. There were appreciable changes in the background characteristics studied for all three groups. None of them made a major contribution to accounting for the score decline among high school seniors or college entrants. A decrease in the percentage of SAT takers entering four-year colleges may help to explain the greater score decline observed for this group.

Braswell, J., and Petersen, N. *An Investigation of Item Obsolescence in the Scholastic Aptitude Test.* October 1976 (revised January 1977), 74 pages.

Two panels, one concentrating on the verbal sections of the SAT and the other on the mathematical, were appointed to review and rate questions that appeared in earlier editions of the test and then in more recent editions. Raters were asked to indicate how the difficulty of each question might be expected to change between the two administration dates. The raters' predictions were then compared with available item analysis data. For the most part, the changes predicted by raters were not substantiated by the statistical analysis. While the relative difficulty of some questions changed between administration dates, it was not possible, except for a few mathematical questions, to attribute these changes to curricular change or to broader social factors. These mathematical questions were predicted by the raters to be relatively easier at the more recent administration, and their prediction was supported by the statistical analysis.

Breland, H. M. *Family Configuration Effects and the Decline in College Admissions Test Scores: A Review of the Zajonc Hypothesis.* September 1976 (revised February 1977), 15 pages.

A hypothesis that part of the SAT score decline is a result of changing American family sizes and configurations is explored. This possible explanation of declining SAT scores had been offered by Robert B. Zajonc in an article in *Science* that reviewed the evidence for a relation between family configuration and cognitive development. Since a number of investigations have shown that "earlyborn" students—those who were the first or second child in their families—and members of small families tend to have higher scores on tests such as the SAT, a change in the representation of students who are earlyborn and members of small families in the population could possibly cause a decline in the average for the total population. The hypothesis is explored through a consideration of the magnitude of the change in family configurations over the years of interest and the magnitude of observed score differences for the SAT. It is concluded that, while the Zajonc hypothesis seems sound, it could only account for a small portion of the total SAT score decline.

Breland, H. M. *The SAT Score Decline: A Summary of Related Research.* January 1976, 40 pages.

This paper surveys the available evidence pertaining to the score decline in terms of five hypothesis areas: the test, the test-taking population, the college-bound population, the schools, and societal factors. It contains a description of the SAT score decline in both graphic and tabular form as well as population data for the years 1957 through 1973. Based on the data available at the time the panel began its deliberations, it is concluded that the evidence confirms a general decline in the abilities of the college-bound population but is less clear with respect to the high school population. It is speculated that the declines in the college-bound population are due to several factors acting in combination: increases in the proportion of low- and middle-income students, reductions in the numbers of test repeaters, changes in the mix of SAT-user colleges, and less-strict policies regarding college admissions.

Chall, J. S., with Conrad, S. S., and Harris, S. H. *An Analysis of Textbooks in Relation to Declining SAT Scores.* November 1976 (revised March 1977), 99 pages.

The authors analyzed the reading selections from six SATS, two from the stable years (1947 and 1955), one from a pivotal year (1962), and three from the declining years (1967, 1972, 1975). Samples of the textbooks used most widely by these six SAT cohorts during their elementary and high school years were analyzed – 35 textbooks and approximately 20 workbooks and teacher's guidebooks in reading, grammar and composition, literature, and history for grades 1, 6, and 11.

The various textbooks and SAT reading passages were analyzed using various indices of difficulty and challenge – the Dale-Chall Readability Formula for level of reading/comprehension difficulty, Chall's Reading Stages for level of linguistic and cognitive maturity, a rating scale for assessing question difficulty adapted from Bloom's Taxonomy, etc.

1. Readability analyses of the SAT passages reveals a general decreasing trend in difficulty from the two stable years (1947 and 1955) through the pivotal year (1962), to the declining years (1967, 1972, and 1975).

2. A possible gap between the difficulty of the SAT passages and the difficulty of the 11th-grade textbooks is indicated. The SAT passages proved to be the most difficult of any of the materials analyzed – more difficult than any of the textbooks. Two of the SAT passages were on the level of grades 13-15 and four on the level of grades 11-12. The history, literature, and grammar and composition 11th-grade textbooks had an average readability level of grades 9-10.

3. There are signs in the data of a recent increasing challenge in the textbooks, particularly at the elementary level.

4. The authors find what appears to be a particularly low level of challenge in writing. Generally the assignments in reading, history, and literature textbooks ask mostly for underlining, circling, and filling in of single words. Few assignments ask students to write a paragraph, story, letter, or theme.

Cleary, T. A., and McCandless, S. A. *Summary of Score Changes (in Other Tests).* February 1, 1977, 10 pages. An excerpt from the authors' "Score Declines and Grade Inflation," a paper delivered at the annual meeting of the Northwest Association

of Schools and Colleges, Portland, Oregon, December 1976.
Summarizes the experience of other testing programs, indicating the degrees to which average scores in each changed over the period of the SAT score decline.

Donlon, T., and Echternacht, G. *A Feasibility Study of the SAT Performance of High-Ability Students from 1960 to 1974 (Valedictorian Study).* October 1976 (revised February 1977). 53 pages.
This is a study of the SAT records of valedictorians and salutatorians in three groups of schools: "experimental" schools, specifically chosen for demographic stability; "comparison" schools, selected for stability but by less stringent criteria; and "private" schools. For each group SAT-V and SAT-M scores in the even-numbered years from 1960 to 1974 were studied.

For the "experimental" group, there is no evidence of score decline in either SAT-V or M. Instead the trends have been toward very modest annual increases. For the "comparison" group the data indicated an initial period of increase, followed by a decrease. Each of these two special samples differed significantly from the SAT-taking population. The "private" schools, however, were not significantly different, showing a decline similar to the total College Board population.

Echternacht, G. *A Comparative Study of Secondary Schools with Different Score Patterns.* October 1976 (revised January 1977). 52 pages.
In this study, the curriculum, institutional, teacher, and student factors associated with those schools having large decreases in SAT score averages were compared with the same factors associated with schools having increasing or steady SAT score averages. It was believed that by identifying these factors, some insight into the role that school characteristics have played in the score decline might come to light. Although some significant differences between the two groups were found in terms of enrollments in academic courses, ability grouping, age of schools, dropout rate, and teacher experience, these were judged not to explain a large portion of the decline. Indeed, differences among schools do not appear to have had large differential effects on the decline of scores.

Ford, S. F., and Campos, S. *Summary of Validity Data from the Admissions Testing Program Validity Study Service.* June 1977.
Validity data (prediction of first-year grade-point average) for colleges participating in the Admissions Testing Program Validity Study Service (VSS) and based on students entering college in 1964 through 1974 are summarized for the following predictors: SAT-verbal score, SAT-mathematical score, high school record, and these three predictors combined.

No definite trends over time are apparent for SAT validities, although those obtained for 1973 and 1974 were generally among the highest observed. There was a downward trend in the validity of high school record accompanied by a slight downward trend in multiple correlations for all three predictors combined. Median validities for colleges having SAT-V means between 450 and 549 tended to be higher than those for colleges having means below 450 or above 550. The great majority of colleges participating in the VSS were found to be four-year colleges. Separate analyses of the small number of two-year colleges in the sample yielded slightly lower median validities for each of the predictors than those found for the total sample.

Glover, R. H. *Major Societal Changes in U. S. (1933-44 and 1945-75): Contextual Mapping.* April 1976 (revised August 1976).

For a span of 42 years (1933-75) a fold-out chart displays major events and influences on events, products, and other phenomena, suggesting their interconnectedness in graphic form. The presentation covers occurrences in the following areas: political, social and cultural, economic and financial, legislative and legal, population and human ecology, and knowledge and technology. The August 1976 revision is accompanied by a scenario for the period 1976-85.

Jackson, R. *Comparison of SAT Score Trends in Selected Schools Judged To Have Traditional or Experimental Orientations.* October 1976, 7 pages.

A group of schools consisting largely of highly regarded schools in affluent suburban areas was divided into two groups—those judged to have a traditional orientation and those judged to have a more experimental orientation (in terms of course and program structure). A review of the mean SAT scores of SAT candidates from the 1966, 1969, 1973, and 1976 graduating classes from these schools shows that (a) mean SAT scores for both school groups were substantially higher than national averages, and (b) declines in mean SAT scores for both groups over the period studied very nearly paralleled the national declines.

This was a small-scale pilot test using available data. Because of the relatively fallible procedures used for classifying schools and because of a general lack of experimental controls of the effects of extraneous factors, the results cannot support any general conclusions about the possible relation of experimentation in the schools to score decline. Because of the essentially negative findings of this pilot test, a more substantial study of this group of schools was not attempted.

Jackson, R. *Correlations of SAT Scores with High School Record.* January 1977, 6 pages.

Two sets of data are examined: (a) correlations of SAT scores with student reports of class standing (collected by means of the Student Descriptive Questionnaire) for 1971-72 through 1975-76, and (b) correlations of SAT scores with measures of high school performance for groups of students attending colleges participating in the College Board Validity Study Service, for entering classes from 1964 through 1974. The correlations of SAT scores with self-reported class rank are virtually level over the five years studied. The longer-term comparisons using validity study data suggest that the median correlations of SAT verbal or mathematical scores with high school record may have increased somewhat over the 11 years studied.

Jackson, R. *An Examination of Declining Numbers of High-Scoring SAT Candidates.* January 1977, 12 pages.

The sharp declines in numbers of SAT candidates scoring over 600, which were observed from 1969-70 through 1974-75, are examined. Several suggested explanations of this phenomenon relating to test-taking patterns are discussed. Specifically examined are the possibilities that less SAT-to-SAT repetition, or less PSAT-to-SAT repetition, or less overlap between SAT and ACT among high-scoring students might account for some part of the observed declines. Although ideal data for testing all these explanations were not available, certain data in hand suggest that, while changes in test-taking patterns may have had some effect, they probably cannot entirely explain the trend in question.

Jackson, R. *Mean Scores for PSAT and PSAT/NMSQT Junior Candidates.* December 20, 1976, 5 pages.
Gives the numbers of students taking the Preliminary Scholastic Aptitude Test and the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test and their mean scores for the period 1959 through 1976. The implications of these results in relation to changes in the candidate group are discussed.

Jackson, R. *A Summary of SAT Score Statistics for College Board Candidates.* December 1975 (revised October 1976), 20 pages.
Summaries of SAT score statistics for test candidates are presented for students grouped according to testing years (for 1956-57 to 1975-76) and for students grouped according to high school classes (from 1966-67 to 1975-76). For recent years, mean scores for geographical regions and for students grouped according to selected responses to the Student Descriptive Questionnaire are also presented.

Jackson, R. *Trends in Mean SAT Scores for Students from the High School Classes of 1971-72 and 1975-76 Belonging to Selected School and College Groups.* January 1977, 9 pages.
Data are presented on trends in SAT performance from the year 1971-72 to 1975-76 for students who (a) attended certain selected high schools for which group summary reports had been produced through the College Board summary report service, or (b) had their scores reported to colleges in groups similarly formed for summary reporting purposes. All school and college groups for which reports were produced both in 1971-72 and 1975-76 are included. The great majority of these groups exhibited declines in mean SAT scores over this period. With only a few exceptions, those groups showing increases in either SAT-verbal or SAT-mathematical mean scores had substantially smaller numbers of students in the more recent year.

Modu, C., and Stern, J. *The Stability of the SAT-Verbal Score Scale.* September 1976, 35 pages.
This study was designed to assess the stability of the score scale between 1963 and 1973 for the verbal sections of the SAT. In a previous study, scores on two old forms were equated to a 1973 form and, through that form, to the College Board scale by means of sets of items common to the new and old forms. This earlier study suggested that the SAT scale had shifted upward by an average of 14 points on the verbal sections and 17 points on the mathematical sections.

In the present study, pairs of 1963 and 1973 forms of the SAT-V were administered in counterbalanced order to spaced samples of the same group, with each candidate taking a 1963 and a 1973 form. The obtained scores were used to place the 1973 scores on the reporting scale used for the 1963 form. The experimentally derived scores on the 1963 scale were then compared with their corresponding scores on the 1973 scale for candidates of the same ability levels in order to estimate the degree of scale shift.

The findings of the present study confirm an upward scale drift which gave the 1973 candidate group an average of 8 to 10 points higher than they would have earned had these experimental equating results rather than the operational equating results been used in reporting the 1973 scores. The scale drift observed for the present study was not uniform over the extent of the scale; it was found to increase as scores decreased from 600 to 200. Little scale drift was noticed over the 10-year span for scores of 650 and above. Based on the evidence from this study it would appear that the reported declines in mean SAT-verbal scores from 1963 to 1973 are about 8 to 10 points smaller than they would have been had the scale been completely stable.

Nelson, J. E. *A Review of Data Available regarding Family Income and Financial Aid Characteristics of Students.* October 1976, 7 pages.
Greatly expanded programs of financial aid based on need have made it possible for large numbers of low-income students to attend college. Since test scores have high correlation with family income, has the increase in low-income students taking the SAT produced the decline in mean scores? This review of available College Board candidate data, income distributions of entering freshmen, percentage of age groups in college by income levels, and impact of student aid on low-income enrollment shows an increased percentage of students coming from lower-income families before 1972, but no significant change since that time.

Reed, J. S. *Available Evidence on Public Attitudes toward Education.* June 1976 (Draft), 31 pages.
Questions dealing with education asked by American survey organizations since 1936 are examined, and over 60 taken to illustrate (1) changes in public opinion since 1960 that might have altered students' evaluations of college going or academic excellence, and (2) changes prior to 1960 that might have affected the quality of preparation for college.

Tables summarize responses to the following broad groups of questions: satisfaction with public schools and personnel and with college and university personnel; adequacy of homework in the public schools, of time spent in school, of discipline in the public schools, of the public school curriculum, of treatment of students with differing ability; desirability of nationwide testing programs; and evaluation of the teaching profession.

It is found that there was a period of criticism of all levels of schooling during the late 1950s, increasing satisfaction with schools and colleges during the 1960s, and a renewed dissatisfaction that starts at the end of the sixties. It would probably be impossible to reconstruct opinion changes between 1950 and 1970 in sufficient detail to relate them to changes in SAT scores.

A catalog of survey questions dealing with education, other than the questions analyzed in the report, is appended.

——A supplement to Reed's paper, consisting of excerpts from the *Eighth Annual Survey of the Public's Attitudes toward the Public Schools.* A project jointly conducted by The Gallup Poll and the Charles F. Kettering Foundation. Princeton, New Jersey: Public Opinion Surveys, Spring 1976.

A significant drop in the public's rating of the schools was recorded between 1974 and 1975, but this decline leveled off between 1975 and 1976. Adults perceive the following as the major problems of the local public schools: lack of discipline; integration, segregation, and busing; lack of proper financial support; poor curriculum; use of drugs; difficulty of getting "good" teachers; parents' lack of interest; size of school or classes or both; school board policies; pupils' lack of interest.

The public believes that the decline in national test scores in recent years means that the quality of education is declining; it would like more attention paid to basic skills and to discipline in the schools in order to counteract the perceived decline of quality in the schools. It inclines, however, to place blame on parents rather than preponderantly on the schools.

Rosenthal, E., and Beaton, A. E. *Annotated Bibliography.* December 1, 1975 (revised April 28, 1976), 17 pages.

More than 30 reports and publications pertaining to national declines in abilities are cited, and most are annotated. The citations in this bibliography include evidence available at the time the panel began its work.

Schrader, W. B. *Distribution of SAT Scores to Colleges as an Indicator of Changes in the SAT Candidate Population.* September 1976, 14 pages.

This study identifies trends in the extent to which SAT candidates were applying to various groups of colleges in 1960-61, 1966-67, and 1973-74. Colleges were grouped on the basis of: (1) the classification developed by the Carnegie Commission on Higher Education, (2) the test or tests they required applicants to take, and (3) state and region. Results provide pertinent data on two main topics: (a) widening access to higher education and (b) acceptance of either SAT or ACT scores for admissions by some colleges. Although the interpretation of the data is complicated, especially by the lack of knowledge about the relation between college choice and taking the SAT, the results offer some reason to believe that both these factors resulted in changes in test taking between 1966-67 and 1971-72 that could have contributed to the SAT score decline. They do not, however, provide a numerical estimate of the size of the effects.

Schramm, W. *Television and the Test Scores.* August 1976, 31 pages.

There is no conclusive evidence as yet that television has been a sufficient cause for decline in test scores, although it may be one of several elements in a complex causal system. The major studies show that television viewing, after the early school years, tends to be associated with lower-than-average achievement, although the relative extent to which viewing affects achievement or unsatisfactory achievement encourages children to take refuge in television is not fully understood. Television reduces reading time, social interaction time, and the opportunity to practice certain skills necessary to academic excellence. It tends to reduce the average level of intellectual stimulation available to a child after the age of 9 or so. The trend of the evidence is that television viewing patterns belong to a group of strong variables that interact with each other and with school (and, therefore, test) performance, probably with negative effect.

Stern, J. *Selected Percentiles for Scholastic Aptitude Test Scores (1966-67 through 1975-76).* January 1977, 1 page.

This table presents the ninetieth, seventy-fifth, fiftieth, twenty-fifth, and tenth percentile groupings for SAT Verbal and Mathematical scores reported in the 10-year period between 1966-67 and 1975-76. A testing year is defined as extending from September to August; data are collected without regard to level of preparation or the number of times an individual student was tested during the defined year.

Stern, J. *Table of SAT and Achievement Test Scores for Samples of Candidates Taking Achievement Tests 1966-67 to 1975-76.* January 1977, 2 pages.

The table provides the SAT and the Achievement Test means and standard deviations for samples of candidates taking each of the 15 Achievement Tests offered in the Admissions Testing Program battery during the period 1966-67 to 1975-76. The samples, drawn annually through 1971-72 and biennially thereafter, provide estimates of SAT means for each Achievement Test population.

Wharton, Y. *List of Hypotheses Advanced to Explain the SAT Score Decline.* December 3, 1975 (revised February 20, 1976), 8 pages.

A listing of hypotheses to explain the SAT score decline (advanced in letters to the College Board, to the Advisory Panel, in magazine articles, and in newspaper stories) is presented and categorized using a four-part classification scheme: changes in the schools, changes in society, changes in the population, and problems with the tests. The first major category (changes in the schools) is further broken down into hypotheses relating to curriculum, institutional policies, teachers, and students. The second major category (changes in society) lists hypotheses related to family, religion, civil rights, crisis of values, national priorities, economic, labor movement in education, and technological changes.

Winter, D. G. *Motivational Factors in the SAT Score Decline.* October 1976 (Draft), 23 pages.

The paper examines the research evidence and theory about three human social motives that could be expected to play some role in test performance and academic functioning: the motives for achievement, for affiliation, and for power. It also considers the possibility of a "motivational overload." Broad cultural trends (i.e., post-Sputnik emphasis on academic achievement, revolt of the counterculture, crisis of values in Vietnam and Watergate) and related specific changes in the educational system ("open," "alternative," and "humanistic" education) may have had direct and indirect effects on motives. Some of the hypotheses suggest that actual motive levels went up or down with corresponding effects on academic performance. Others suggest the ways in which social forces may have changed beliefs about the nature and value of academic work, beliefs that may interact with relatively constant motive levels to produce behavior changes.

Bibliography

Commissioned by the Panel

Arnold, C. B., *Could There Be a Medical Basis for the Declining SAT Scores?* New York: College Entrance Examination Board, 1977.

Beaton, A. E., Hilton, T. L., and Schrader, W. B., *Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates between 1960 and 1972.* New York: College Entrance Examination Board, 1977.

Braswell, J., and Petersen, N., *An Investigation of Item Obsolescence in the Scholastic Aptitude Test.* New York: College Entrance Examination Board, 1977.

Breland, H. M., *Family Configuration Effects and the Decline in College Admissions Test Scores: A Review of the Zajonc Hypothesis.* New York: College Entrance Examination Board, 1977.

Breland, H. M., *The SAT Score Decline: A Summary of Related Research.* New York: College Entrance Examination Board, 1977.

Chall, J. S., with Conrad, S. S., and Harris, S. H., *An Analysis of Textbooks in Relation to Declining SAT Scores.* New York: College Entrance Examination Board, 1977.

Cleary, T. A., and McCandless, S. Q., *Summary of Score Changes (in Other Tests).* New York: College Entrance Examination Board, 1977.

Donlon, T., and Echternacht, G., *A Feasibility Study of the SAT Performance of High-Ability Students from 1960 to 1974 (Valedictorian Study).* New York: College Entrance Examination Board, 1977.

Echternacht, G., *A Comparative Study of Secondary Schools with Different Score Patterns.* New York: College Entrance Examination Board, 1977.

Ford, S. F., and Campos, S., *Summary of Validity Data from the Admissions Testing Program Validity Study Service.* New York: College Entrance Examination Board, 1977.

Glover, R. H., *Major Societal Changes in U.S. (1933-44 and 1945-1975): Contextual Mapping.* New York: College Entrance Examination Board, 1977.

Jackson, R., *Comparison of SAT Score Trends in Selected Schools Judged To Have Traditional or Experimental Orientations.* New York: College Entrance Examination Board, 1977.

Jackson, R., *Correlations of SAT Scores with High School Record.* New York: College Entrance Examination Board, 1977.

Jackson, R., *An Examination of Declining Numbers of High-Scoring SAT Candidates.* New York: College Entrance Examination Board, 1977.

Jackson, R., *Mean Scores for PSAT and PSAT/NMSQT Junior Candidates.* New York: College Entrance Examination Board, 1977.

Jackson, R., *A Summary of SAT Score Statistics for College Board Candidates.* New York: College Entrance Examination Board, 1977.

Jackson, R., *Trends in Mean SAT Scores for Students from the High School Classes of 1971-72 and 1975-76 Belonging to Selected School and College Groups.* New York: College Entrance Examination Board, 1977.

Modu, C., and Stern, J., *The Stability of the SAT-Verbal Score Scale.* New York: College Entrance Examination Board, 1977.

Nelson, J. E., *A Review of Data Available regarding Family Income and Financial Aid Characteristics of Students.* New York: College Entrance Examination Board, 1977.

Reed, J. S., *Available Evidence on Public Attitudes toward Education.* New York: College Entrance Examination Board, 1977.

Rosenthal, E., and Beaton, A. E., *Annotated Bibliography.* New York: College Entrance Examination Board, 1977.

Schrader, W. B., *Distribution of SAT Scores to Colleges as an Indicator of Changes in the SAT Candidate Population.* New York: College Entrance Examination Board, 1977.

Schramm, W., *Television and the Test Scores.* New York: College Entrance Examination Board, 1977.

Stern, J., *Selected Percentiles for Scholastic Aptitude Test Scores (1966-67 through 1975-76).* New York: College Entrance Examination Board, 1977.

Stern, J., *Table of SAT and Achievement Test Scores for Samples of Candidates Taking Achievement Tests 1966-67 to 1975-76.* New York: College Entrance Examination Board, 1977.

Wharton, Y., *List of Hypotheses Advanced to Explain the SAT Score Decline.* New York: College Entrance Examination Board, 1977.

Winter, D. G., *Motivational Factors in the SAT Score Decline.* New York: College Entrance Examination Board, 1977.

General

Abt Associates, *Education as Experimentation.* Report prepared for the Office of Education. Washington, D.C.: United States Department of Health, Education, and Welfare, 1977, Vol. IV-A.

American College Testing Program, *The High School Profile Report, Years 1975-76.* Iowa City Iowa: American College Testing Program, 1977.

American Council on Education, "National Norms for Entering College Freshmen – Fall 1966." *ACE Research Reports,* 1967, Vol. 2, No. 1.

American Council on Education, "The American Freshman: National Norms for Fall 1972." *ACE Research Reports,* 1972, Vol. 7, No. 5.

American Federation of Teachers, "To Test or Not To Test: An Examination of the Educational Testing Controversy." *American Educator,* Winter 1977.

Anchor Test Study, Fifth-Grade Reading Achievement as a Function of Selected School, Classroom, and Pupil Variables. Washington, D.C.: National Center for Educational Statistics, 1976.

Angoff, William H. (editor), *The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests.* New York: College Entrance Examination Board, 1971.

Angoff, W., "Why the SAT Scores Are Going Down." *English Journal,* March 1975, pp. 10-11.

Armbruster, Frank E., *The U. S. Primary and Secondary Educational Process.* With contributions by Paul J. Bracken and the assistance of Joyce Lind. Final Draft. Croton-on-Hudson, N.Y.: Hudson Institute, 1975.

Astin, Alexander et al., *The American Freshman: National Norms* (for years 1966-76). Cooperative Institutional Research Program, American Council on Education and University of California at Los Angeles.

Bannister, Robert C. (Professor of History, Swarthmore College, Swarthmore, Pennsylvania), Letter to George H. Hanford, College Entrance Examination Board, October 8, 1976.

Beshoar, Barron B., "The Condition of Student Writing." *American Education,* March 1976.

Brewster, Galen, "Anxiety, Identity, and Declining Scores." *Independent School,* October 1976, pp. 15-18.

Bullock, R., and Stern, J., *Estimated SAT Summary Statistics for High School Cohorts 1967-1974.*

Memorandum dated September 3, 1974. Princeton, N.J.: Educational Testing Service.

Buros, Oscar K. (editor), *The Seventh Mental Measurements Yearbook.* Reviews of the College Board Scholastic Aptitude Test by Philip H. DuBois and Wimburn L. Wallace. New Jersey: The Gryphon Press, 1972, Vol. 1, pp. 640, 646-650.

A Century of Public School Statistics (1975). Washington, D.C.: National Center for Educational Statistics.

Coleman, James S., *Equality of Educational Opportunity.* Washington, D.C.: Office of Education, United States Department of Health, Education, and Welfare, 1966.

College-Bound Seniors, 1974-75. New York: College Entrance Examination Board, 1975.

College-Bound Seniors, 1975-76: National Report. New York: College Entrance Examination Board, 1976.

The Condition of Education (1975-77). Washington, D.C.: National Center for Educational Statistics.

The Decline in Achievement in Science and Other Disciplines. Edison Foundation and the Institute for Development of Educational Activities, Inc., 1975.

Declining Test Scores: A Conference Report (edited by Evelyn Stein Silver, Lawrence Johnson & Associates, Inc.). Washington, D.C.: National Institute of Education, February 1976.

Digest of Educational Statistics, 1975 edition. Washington, D.C.: National Center for Educational Statistics, Table 10, p. 14.

Ebel, Robert L., "Declining Scores—Why?" Paper prepared for presentation at the annual meeting of the North Central Association, Chicago, Illinois, on March 30, 1976.

Effects of Coaching on Scholastic Aptitude Test Scores. New York: College Entrance Examination Board, 1965.

Ernest, John, *Mathematics and Sex.* Santa Barbara, California: Mathematics Department, University of California at Santa Barbara, 1976.

Farr, R., Turnman, J., and Rowls, M., *Reading Achievement in the United States: Then and Now.* Report prepared for the United States Office of Education by Educational Testing Service and Indiana University, 1975.

Fasold, J., Hall, M., and Impara, J., *Decline in Standardized Test Scores: A Widespread Phenomenon.* Salem, Oregon: Oregon Department of Education, 1974.

Ferguson, Richard L., and Maxey, James E., *Trends in the Academic Performance of High School and College Students.* American College Testing Program Report No. 70. Iowa City, Iowa: American College Testing Program, January 1976.

Flanagan, John C., "Changes in School Levels of Achievement: Project TALENT Ten and Fifteen Year Retests." Paper presented at the Symposium of Division D and National Council on Measurement in Education during the 1976 meeting of the American Educational Research Association, San Francisco, April 22, 1976.

Flanagan, John C., and Jung, Steven M., *Progress in Education: A Sample Survey* (1960-70). Palo Alto, California: American Institutes for Research, December 1971.

Foster, B. J., *Statistics of Public Elementary and Secondary Day Schools, Fall 1972.* Department of Health, Education, and Welfare Publication No. (OE) 73-11402. Washington, D.C.: United States Government Printing Office, 1973.

Froomkin, Joseph, "Declining Test Scores: Reasons and Impacts." Paper prepared for the Higher Education Policy Research Center, United States Department of Health, Education, and Welfare. Washington, D.C.: United States Government Printing Office, 1976.

Functional Literacy, Basic Reading Performance (1974 and 1975). National Assessment of Education Progress Report. Denver, Colorado: Education Commission of the States.

Gallup, George H., "Eighth Annual Gallup Poll of the Public's Attitudes toward the Public Schools." Reprinted from the October 1976 *Phi Delta Kappan.*

Giamatti, A. Bartlett, "The Language Barrier: Why Students Can't Write." *Yale Alumni Magazine*, January 1976.

Gleazer, Edmund J., Jr., "Declining Achievement of High School Graduates." Presentation to the Cosmos Club, Washington, D.C., January 18, 1977.

Graves, Donald, "Writing: An Endangered Species." *The Common*, March 1977.

Guide for High Schools and Colleges (Interpretive publication of the Admissions Testing Program). New York: College Entrance Examination Board, 1975.

Hanford, George H., "Change in SAT Scores." Remarks prepared for presentation at the 1976 annual meeting of the North Central Association, Chicago, Illinois, on March 30, 1976.

Harnischfeger, A., and Wiley, D., *Achievement Test Score Decline: Do We Need to Worry?* Chicago, Illinois: CEMREL, Inc., 1975.

Hawes, G. R., "The Decline of the SAT's." *CHANGE*, November 1972, pp. 17-19.

Herndon, Terry (executive director, National Education Association), Letter to S. P. Marland, Jr., College Entrance Examination Board, April 9, 1976.

Hilton, T. L., Rhett, H., Broudy, I. L., Bower, C., Carter, M. M., Creech, F. R., and Echternacht, G. J., *The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972.* Final Report. Princeton, N.J.: Educational Testing Service, 1973.

Houts, Paul L. (editor), *The Myth of Measurability*. New York: Hart Publishing Co., 1977.

Howell, Barbara Thompson, Presentation to New Jersey Association of School Administrators, Atlantic City, New Jersey, May 12, 1977.

Impact of Educational Innovation on Student Performance. Project Longstep – Final Report: Vol. I, Executive Summary. Palo Alto, California: American Institutes for Research, 1976.

"Inside the Brain: The Last Great Frontier." Symposium: Albert Rosenfeld and Kenneth Klivington, Maya Pines, Richard Restak, Howard Gardner, Roger W. Sperry. *The Saturday Review*, August 9, 1975.

Jackson, Rex, Paper for the panel showing tabulation of College Board's Student Descriptive Questionnaire data – 1973-77.

Jackson, Rex, and Schrader, William B., *Verbal and Mathematical Ability of High School Juniors in 1974: A Norms Study of PSAT/NMSQT.* College Entrance Examination Board Research and Development Reports RDR-76-77, No. 2. New York: College Entrance Examination Board, 1976.

Jencks, C., *Inequality: A Reassessment of the Effect of Family and Schooling in America.* New York: Basic Books, 1972.

Johnson, Charles E., Jr., "Changing Trends in College Enrollment." *College and University*, Fall 1973, pp. 35-41.

Karpinos, Bernard D., *AFQT [Air Force Qualifying Test]: Historical Data (1958-1972).* Alexandria, Virginia: Manpower Research and Data Analysis Center, Human Resources Research Organization, 1975.

Kifer, Edward, "The Relationship between the Home and School in Influencing the Learning of Children." Paper delivered at the Pre-Convention Conference on Research, National Council of Teachers of English, Chicago, Illinois, November 23, 1976.

Kolata, Gina Bari, "Aftermath of the New Math: Its Originators Defend It." *Science*, March 4, 1977, pp. 854-857.

Layton, Katherine P., Informal report to George H. Hanford on a meeting of the Mathematical Association of America, October 19, 1976.

Maguire, John, "That Which Comes Down, Must It Go Up?: The Decline of SATS." Paper presented at the College Board New England Regional Meeting, February 1976.

Manchester, William, "Up against the Wall." In *The Glory and the Dream.* Boston: Little, Brown and Company, 1974.

Marco, G. L., and Stern, J., *Investigation of the "Repeater" Hypothesis Concerning SAT Score Declines.* Princeton, N.J.: Educational Testing Service, April 1974.

Maxey, E. James, *Trends in the Academic Abilities, Background Characteristics, and Educational and Vocational Plans of College-Bound Students.* American College Testing Program Report No. 74. Iowa City, Iowa: American College Testing Program, May 1976.

McCandless, Sam A., "The SAT Score Decline and Its Implications for College Admissions." Paper presented at the College Board Western Regional Meeting, January 1975.

Modu, Christopher C., and Stern, June, *The Stability of the SAT Score Scale.* Princeton, N.J.: Educational Testing Service, 1975 (RB-75).

Munday, Leo A., *Declining Admissions Test Scores.* ACT Research Report No. 71. Iowa City, Iowa: American College Testing Program, February 1976.

National Assessment of Science (1969-73). National Assessment of Educational Progress Report. Denver, Colorado: Education Commission of the States.

National Association of Secondary-School Principals, "Student Attendance and Absenteeism." *The Practitioner,* March 1975.

National Education Association, "Special Feature on Testing (1977)"—reprints of articles by Edwin F. Taylor, Charlotte Darehshori, and Brenda S. Engel. *Today's Education,* March-April 1977.

"Newest National Norms Show—Decline in Scores Continues." *ACTivity,* October 1975, 13 (4), p. 5.

Overview and Analysis of School Mathematics, Grades K – 12. Washington, D.C.: Conference Board of the Mathematical Sciences, National Advisory Committee on Mathematical Sciences, 1975.

Peng, Samuel S., "Some Trends in the Entry to Higher Education: A Comparison between NLS and Project TALENT." Paper presented at the annual meeting of the American Psychological Association in Washington, D.C., 1976.

Perry, D., and Swanson, E. O., "Decline in Minnesota College Aptitude Test Scores." In J. Fasold et al., eds., *Decline in Standardized Test Scores: A Widespread Phenomenon.* Salem, Oregon: Oregon Department of Education, 1974.

Pike, Lewis W., and Evans, Franklin, R., *Effects of Special Instruction for Three Kinds of Mathematics Aptitude Items.* College Entrance Examination Board Research Report. New York: College Entrance Examination Board, 1972.

Randhawa, Bikkar S., "Influence of Visual Domain on Score Decline: Some Conjectures." Unpublished paper, 1977.

Reading in America: A Perspective on Two Assessments. National Assessment of Educational Progress Report. Denver, Colorado: Education Commission of the States.

Report on a National Study of the Nature and the Quality of Instructional Materials Most Used by Teachers and Learners. Educational Products Information Exchange Report No. 76, 1977.

Schrader, W. B., *Test Data as Social Indicators.* Statistical Report SR-68-77. Princeton, N.J.: Educational Testing Service, September 1968.

Schrader, W. B., and Hilton, T., *Educational Attainment of American High School Seniors in 1960, 1965, and 1972: Feasibility Study.* Final Report. Project No. NIE-G-74-0050, 1975. Princeton, N.J.: Educational Testing Service, 1975.

Schrader, W. B., and Jackson, R., *Preliminary Report on the 1974 PSAT/NMSQT Norms Study.* Memorandum dated March 24, 1975. Princeton, N.J.: Educational Testing Service.

Start, K. B., "Thirty Years of Reading Standards in England." *Educational Researcher,* 1972, 1 (11), pp. 8-9.

Stewart, E. E., *The Stability of the SAT-Verbal Score Scale.* Research Bulletin RB-66-37, CEEB RDR-66-7. Princeton, N.J.: Educational Testing Service, July 1966.

Student Bulletin, 1975-1976, Admissions Testing Program. New York: College Entrance Examination Board, 1975.

Survey of Curricula Offerings. Massachusetts Department of Education, 1967-68, 1970-71, 1972-73, 1975-76.

Tavris, Carol, "The End of the IQ Slump." *Psychology Today,* April 1976, pp. 69-74.

Television and Children, Priorities for Research. Report of Ford Foundation Conference at Reston, Virginia, November 5-7, 1975.

Testing and Evaluation: New Views. Washington, D.C.: Association for Childhood Education International, 1975.

Thorndike, R. L., "Mr. Binet's Test 70 Years Later." *Educational Researcher,* May 1975, 4 (5), pp. 3-7.

Weinman, Janice J., *Declining Test Scores: A State Study.* Massachusetts: Massachusetts Department of Education, 1977.

White, E. B., "One Man's Meat." *Harper's Magazine,* 1938.

Whitney, Hassler, "Learning: Form or Substance?" Paper prepared for the panel by Dr. Hoover of the Institute for Advanced Study, Princeton, New Jersey, May 11, 1976.

Wilks, S. S., *Scaling and Equating College Board Tests.* Princeton, N.J.: Educational Testing Service, 1961.

Wilson, Sandra, and Wise, Lauress, *The American Citizen: 11 Years After High School.* Palo Alto, California: American Institutes for Research, 1975.

Writing Mechanics (1969-74). National Assessment of Educational Progress Report. Denver, Colorado: Education Commission of the States.

Zajonc, R. B., "Family Configuration and Intelligence." *Science,* April 16, 1976, pp. 227-236.

Zajonc, R. B., "Intellectual Environment and Intelligence." Unpublished paper submitted to the panel. Revised October 1975.

Zajonc, R. B., and Markus, G. B., "Birth Order and Intellectual Development." *Psychological Review,* 1975, 82 (1), pp. 74-88.

Inspection copies of the Scholastic Aptitude Test were made available on a confidential basis to members of the Advisory Panel on Score Decline.

News Reports

Barnes, B., "Area Pupils Attend Special Cram Courses to Prep for SAT." *The Washington Post,* October 18, 1976.

"A Big Test for Test Publishers." *Business Week,* June 27, 1977.

Briggs, B. Bruce, "The Great Classroom Debacle." *The Wall Street Journal,* July 20, 1976.

Bronfenbrenner, Urie, "The Calamitous Decline of the American Family." *The Washington Post,* January 2, 1977.

Cattani, Richard J., "U.S. Colleges on 'Back to Basics Trail.'" *The Christian Science Monitor,* May 12, 1977.

Feinberg, Lawrence, "Teacher Standards Issue Upsets City." *The Washington Post,* May 16, 1977.

Feinberg, Lawrence, "'Basic' Teaching Methods More Effective, Study Says." *The Washington Post,* June 20, 1977.

Gunther, Max, "How Television Helps Johnny Read." *TV Guide,* September 4, 1976.

Kelch, Kristen, "High School Blends Fundamentals, New Concepts." *The Christian Science Monitor,* October 20, 1976.

King, Wayne, "Stiffer Teacher Test Upsets Blacks in South Carolina." *The New York Times,* June 24, 1977.

Klaidman, Stephen, "Western High Valedictorian Fails GWU Entrance Criteria." *The Washington Post*, August 15, 1976.

Landers, Jim, and Meyer, J. Stryker, Series of articles on development and administration of the SAT. *The Sunday Times-Advertiser* and *The Evening Times*, Trenton, N.J., March 28-31, 1976.

Little, D. Richard, "Legacy of the '60's – Declining Quality." *The Christian Science Monitor*, January 17, 1977.

Maeroff, G., "The Missing School Yardstick: Measuring Gains and Losses." *The New York Times*, March 17, 1975.

Malcolm, Andrew H., "Japanese Students Tensely Undergoing Annual Exam Hell." *The New York Times*, February 24, 1977.

McCurdy, Jack, and Speich, Don, "The Decline of American Education." Reprinted from *The Los Angeles Times*, August 15-17, 1976.

Nathan, Linda K., "'Improving' Law School Aptitude Is Big Business." *The New York Times*, November 14, 1976.

Parsons, Cynthia, "SAT + ACT = College Tests in a Mess." *The Christian Science Monitor*, March 29, 1976.

Peterson, I., "College Textbooks Being Simplified to Meet Needs of the Poor Reader." *The New York Times*, November 7, 1974.

Raspberry, William, "A Dilution of School Standards." *The Washington Post*, May 16, 1977.

Rhoda, Andreas de, "Student-Achievement Slump – Is It the Fault of Schools?" *The Christian Science Monitor*, July 26, 1976.

Sowell, Thomas, "New Light on Black I.Q." *The New York Times Magazine*, March 27, 1977.

"What TV Does to Kids." *Newsweek*, February 21, 1977.

"Why Johnny Can't Write." *Newsweek*, December 8, 1975.

Sample Questions from the SAT

Verbal Questions

The ability to understand what you read and the extent of your vocabulary are important for successful academic performance in college. The SAT tests these abilities using four types of questions: antonyms and analogies, which yield a vocabulary subscore, and sentence completion and reading comprehension, which yield a reading subscore. The directions that are shown below for each type of question are the same as the ones that appear in the test itself.

Vocabulary

Antonyms (opposites)
Antonyms test the quality and extent of your vocabulary.

Directions: Each question below consists of a word in capital letters, followed by five lettered words or phrases. Choose the word or phrase that is most nearly opposite in meaning to the word in capital letters. Since some of the questions require you to distinguish fine shades of meaning, consider all the choices before deciding which is best.

Example:
GOOD: (A) sour (B) bad (C) red (D) hot
(E) ugly Ⓐ ● Ⓒ Ⓓ Ⓔ

Practice Questions:

1. BABBLE: (A) irrigation (B) pollution
(C) meaningful speech (D) useful object
(E) helpful person

8. RECTITUDE: (A) deliberation (B) laziness
(C) prejudice (D) laxity of morals
(E) weakness of intellect

Analogies
Analogies test your understanding of relationships between words and ideas. You are asked to recognize pairs that are similar or parallel in nature.

Directions: Each question below consists of a related pair of words or phrases, followed by five lettered pairs of words or phrases. Select the lettered pair that best expresses a relationship similar to that expressed in the original pair.

Example:
YAWN:BOREDOM:: (A) dream:sleep
(B) anger:madness
(C) smile:amusement (D) face:expression
(E) impatience:rebellion Ⓐ Ⓑ ● Ⓓ Ⓔ

First, establish the relationship between the two capitalized words, considering for each word only the sense that applies. For example, the fact that you yawn when you're sleepy is irrelevant in the example above. Roughly stated, the relationship between the first two words is "a yawn indicates boredom." The correct answer is (C), because "a smile indicates amusement." If more than one of the choices seems to fit the relationship you have established, restate it more precisely. The correct answer to the sample question is more obvious if the relationship is expressed, as "a yawn is a movement of facial muscles indicating boredom."

Practice Questions:

10. CHOIR:SINGERS:: (A) victory:soldiers
(B) class:teachers (C) crowd:protestors
(D) challenge:duelists (E) orchestra:musicians

16. BALLAST:STABILITY:: (A) menu:appetite
(B) buoy:steering (C) spice:flavor
(D) grade:education (E) eclipse:clarity

Reading Comprehension

Sentence Completion
This type of question asks you to select words or phrases that are consistent in style and logic with other elements in the sentence.

Reprinted from the booklet for students, *About the SAT and the Test of Standard Written English, 1977-78.* (Copyright © 1977 by College Entrance Examination Board. All rights reserved. Sample test questions in the booklet are reprinted by permission of ETS, the copyright owner.)

Directions: Each question below has one or two blank spaces, each blank indicating that something has been omitted. Beneath the sentence are five lettered words or sets of words. Choose the word or set of words that best fits the meaning of the sentence as a whole.

Example:

Although the publicity has been -----, the film itself is intelligent, well-acted, handsomely produced, and altogether -----.

(A) extensive . . arbitrary
(B) tasteless . . respectable
(C) sophisticated . . amateurish
(D) risque . . crude
(E) perfect . . spectacular Ⓐ ● Ⓒ Ⓓ Ⓔ

The correct answer should involve two words that are more or less opposite in meaning, as the word although suggests that the publicity misrepresented the film. Another clue to the correct answer is that the second word should fit in context with the words "intelligent, well-acted, handsomely produced." Choices (A), (D), and (E) are not opposites. Choice (C) can not be the correct answer even though the words in it are nearly opposites, because if the film is intelligent, well-acted, and handsomely produced, it is not amateurish. Also only choice (B), when inserted in the sentence, produces a logical statement.

Practice Questions:

19. Intricately carved and beautifully proportioned, the priceless sculpture was ----- the work of a -----.

(A) understandably . . dilettante
(B) indelibly . . forger
(C) demonstrably . . bungler
(D) unmistakably . . master
(E) paradoxically . . perfectionist

Reading Passages

The reading passages are taken from a variety of fields, and reading comprehension is tested at several levels. Some of the questions test your understanding of what has been stated directly; others test your ability to interpret and analyze what you have read. Be sure to read the questions carefully so that you understand exactly what is being asked.

Directions: Read the following passage and then answer the questions on the basis of what is stated or implied in the passage.

The behavioral sciences are making rapid strides in the understanding, prediction, and control of behavior. In important ways we know how to select individuals who will exhibit certain behaviors and to establish conditions in groups which will lead to various predictable group behaviors; in animals our ability to understand, predict, and control goes even further, possibly foreshadowing future steps in relation to man.

If your reaction is the same as mine, then you will have found that the potentials of this young science are somewhat frightening. For all its present immaturity, behavioral science may contain awesome possibilities. If some individual or group had the power to exploit this science, it would be a nightmare of manipulation. Potential troublemakers could be discovered and dealt with before they became such. Morale could be improved or lowered and behavior could be influenced by appeals to motives of which the individual was unconscious. Admittedly this is wild fantasy, but it is not an impossible fantasy.

Some of you may point out that only a few of the findings I have mentioned have actually been put to use in any way that significantly affects society, and that for the most part these studies are important only to the behavioral scientist but have no serious impact on our culture. I agree with this point. The behavioral sciences at the present time are at somewhat the same stage as the physical sciences were several generations ago. For instance in 1900, the public believed the science of aeronautics to be of little importance and did not anticipate the significant effects that aeronautics would have on culture. They preferred to use their common sense, which told them that man could not possibly fly in a contraption which was heavier than air.

However, the public attitude toward physical science is quite different today. The public is ready to believe any prediction the physical scientist might make. When science predicted a satellite would be launched into space, very few voices were raised in disbelief.

There is every reason to believe that the same sequence of events will occur in the behavioral sciences. First, the public ignores or views with disbelief; then, as it discovers that the findings of a science are more dependable than theories based on common sense, it begins to use them; eventually, the widespread use of these findings creates a tremendous demand. Finally, the development of the science spirals upward at an ever-increasing rate. Consequently, even though the findings of the behavioral sciences are not widely used today, there is every likelihood that they will be widely used tomorrow.

Practice Questions:

26. **The author suggests that the next change in the public's attitude toward behavioral science will lead the public to**
    (A) ignore the findings
    (B) increase the use of the findings
    (C) disbelieve the findings
    (D) use these findings against each other
    (E) loose interest in the findings

27. **The tone of this passage can best be described as**
    (A) condescending (B) humble (C) insipid
    (D) admonitory (E) inspiring

Mathematical Questions

The subject matter prerequisites for the SAT-mathematical questions include arithmetic and do not extend beyond a year of high school algebra and the geometry that is usually taught in the elementary and junior high years. The arithmetic includes the four basic operations of addition, subtraction, multiplication, and division; properties of odd and even integers; percent; and averages. The algebra includes linear equations, simple quadratic equations, factoring, and exponents, but not the quadratic formula, fractional or negative exponents, or logarithms. The geometry includes the properties associated with parallel lines and the informal measurement-related concepts of area, perimeter, volume, the Pythagorean Theorem, and angle measure in degrees. Knowledge of special triangles such as isosceles, equilateral, 30°-60°-90° is also assumed. (See question 20.) Unusual notation is used only when it is explicitly defined for a particular question. (See question 17.)

Certain questions emphasize nonroutine problem-solving approaches. For example, the correct solution to problem 15 . . . can be easily obtained once an appropriate nonroutine approach is discovered, but could be obtained only at the cost of much time and thought if a routine approach were used. When you take the SAT, however, do not spend too much time searching for a nonroutine solution. If you cannot think of such a solution, try a routine approach or go on to the next question.

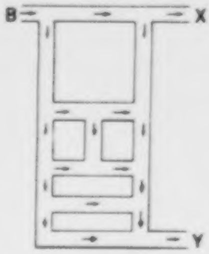
Two kinds of multiple-choice questions appear in the mathematical portion of the SAT:

1. Standard multiple-choice questions (approximately two-thirds of the test)
2. Quantitative comparison questions (approximately one-third of the test)

Each of the quantitative comparison questions presents two quantities to be compared, one in Column A and the other in Column B. . . . . In general, this type of question takes less time to answer, involves less reading, and requires somewhat less computation than the usual multiple-choice mathematics question. In deciding on your answer, you must use concepts of greater than (>), less than (<), and equal to (=) to decide which choice is correct. Please analyze carefully example 27. . . . as this illustrates comparisons in which the relationship cannot be determined. . . .

Practice Questions:

1. **The number which is one hundred less than one million is**
   (A) 990,900 (B) 900,900 (C) 999,990
   (D) 999,900 (E) 1,000,900



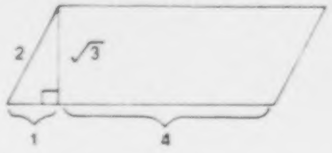
15. **The figure above represents a network of one-way traffic lanes. If the traffic divides equally at intersections where there are alternative directions and in one hour 512 cars enter the traffic pattern at point B, how many of these cars will leave via Y?**
    (A) 128 (B) 192 (C) 256 (D) 320 (E) 384

Question 17 refers to the following definition.

For all positive integers x, $\boxed{x} = x$ **If x is even**

$\boxed{x} = x + 1$ **If x is odd**

17. $\boxed{19} + \boxed{21} =$
    (A) 38 (B) 39 (C) 40 (D) 41 (D) 42



20. **The figure above shows a piece of paper in the shape of a parallelogram with measurements as indicated. If the paper is tacked at its center to a large flat surface and then rotated about its center, the points covered by the paper will be a circular region of diameter**
    (A) $\sqrt{3}$ **(B) 2 (C) 5 (D)** $\sqrt{28}$ **(E)** $\sqrt{39}$

Directions: Each of the following questions consists of two quantities, one in Column A and one in Column B. You are to compare the two quantities and on the answer sheet blacken space

A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

Notes:

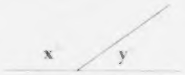
1. In certain questions, information concerning one or both of the quantities to be compared is centered above the two columns.

2. A symbol that appears in both columns represents the same thing in Column A as it does in Column B.

3. Letters such as x, n, and k stand for real numbers.

4. Since there are only four choices, NEVER MARK (E).

Examples:

	Column A	Column B	Answers
E1.	2×6	$2 + 6$	● Ⓑ Ⓒ Ⓓ Ⓔ
E2.	 $180 - x$	y	Ⓐ Ⓑ ● Ⓓ Ⓔ
E3.	$p - q$	$q - p$	Ⓐ Ⓑ Ⓒ ● Ⓔ

	Column A		Column B
		$5x + 3y = 22$	
27.	x		y

Answer Key

Question number	Correct answer	Percentage of 450-scorers who answered correctly
Verbal		
1	C	96
8	D	23
10	E	93
16	C	39
19	D	93
26	B	80
27	D	20
Mathematical		
1	D	90
15	E	22
17	E	47
20	E	20
27	D	39

Advisory Panel on the Scholastic Aptitude Test Score Decline

Willard Wirtz, Chairman, National Manpower Institute, Washington, D.C., *chairman*

Harold Howe II, Vice President, Education and Research, The Ford Foundation, New York, *vice chairman*

Bernard C. Watson, Vice President for Academic Services, Temple University

Ralph W. Tyler, Director Emeritus, Center for Advanced Study in the Behavioral Sciences, Chicago, Illinois

Ledyard R Tucker, Professor of Psychology, Educational Psychology, and in the Center for Advanced Study, University of Illinois at Urbana-Champaign

Vivian H. T. Tom, Teacher of Social Studies, Lincoln High School, Yonkers, New York

Robert L. Thorndike, Professor of Psychology and Education, Teachers College, Columbia University

Barbara Thompson, Superintendent of Public Instruction, State Department of Public Instruction, Madison, Wisconsin

Thomas W. F. Stroud, Associate Professor, Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada

Rosedith Sitgreaves, Professor of Education and Statistics, School of Education, Stanford University

Wilbur Schramm, Distinguished Center Researcher, The East-West Center, Honolulu, Hawaii

Katherine P. Layton, Chairman, Department of Mathematics, Beverly Hills High School, California

Owen B. Kiernan, Executive Secretary, National Association of Secondary-School Principals

H. Thomas James, President and Director, The Spencer Foundation, Chicago, Illinois

Matina S. Horner, President and Dean, Radcliffe College

Edythe J. Gaines, Superintendent of Schools, Hartford Board of Education, Connecticut

Frank W. Erwin, President, Richardson, Bellows, Henry & Company, Inc., Washington, D.C.

Bruce K. Eckland, Professor of Sociology, University of North Carolina at Chapel Hill

Luis C. Cortes, Principal, Bowie High School, El Paso, Texas

Sandra A. Clark, English Department Head, Sammamish High School, Bellevue, Washington

Benjamin S. Bloom, Charles H. Swift Distinguished Service Professor of Education, University of Chicago

Consultants

Edmund J. Farrell, Associate Executive Secretary, National Council of Teachers of English (chairman, College Board English Discipline Committee)

Alfred L. Putnam, Professor of Mathematics, University of Chicago (chairman, College Board Mathematics Discipline Committee)

On Further Examination

Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline

College Board

Could There Be a Medical Basis for the Declining SAT Scores?

Charles B. Arnold
Director, Health Maintenance Institute
American Health Foundation

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

January 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

The problem. The Scores on the Scholastic Aptitude Test (SAT) of the College Entrance Examination Board have been declining since 1963. The total decline has been about 16 percent (to 1975). While this decline has occurred, scores on achievement tests given students from grades 3 to 11 have been stable. The question posed was:

Could there be a health factor that might have caused this situation?

The author's initial reaction was to doubt that there could be, but because the College Board was completing a substantial investigation of the problem, it seemed desirable to have an objective appraisal from a health perspective. The following analysis was completed in approximately six weeks.

Methodology. Documents were provided by the College Board that summarized reports of other task forces. Only one specific question on health was posed: "Could an increased use of induced labor in pregnancy have caused sufficient brain damage among the test-taking population to be a factor in the declining SAT scores?"

We consulted with others and reviewed the medical and epidemiological literature including that on nutrition, drugs, and environmental factors that could have affected infants, children, and adolescents. The technical appendix and bibliography contain a summary of our findings and a list of some of the trends noted in five developmental and related areas.

We particularly examined secular trends in child-development, nutritional patterns, health status, and health care, but found gaps and definitional problems in the published data that necessitated estimations and adjustments in national data. Health care and health status data had not been collected and published in a systematic manner in the United States before 1958, and since then only in specific categories defined by the then-created National Center for Health Statistics (NCHS). In general, the NCHS study categories did not relate to child development issues until the early 1960s and when they did, the critical time period (1945-62) was not investigated in a manner very useful for this report. In the six weeks' period available for our investigation, a picture was imperfectly developed. It is unlikely, however, that a longer time would have produced more. When we conclude, however, that no ecological factor could account for the SAT score decline, there is a certain sense of viewing the problem through a glass darkly.

Before beginning the investigation, a frame of reference was developed that considered the biological reasoning used to determine the direction of the investigation.

Biological reasoning. Four conditions seemed necessary and sufficient before a causal hypothesis regarding the declining SAT scores could be considered seriously. Whatever the nature of a putative health condition, it most likely:

1. Would have occurred in childhood, and have been subclinical in type. If not, it would have been described in the clinical medical literature and recognized as a "medical condition" previously.
2. Would be widely distributed in the United States population because of the numerical size of the test group.
3. Would have begun around 1945-46 since most test takers are 17-18 years old, and the first observed decline was apparent after 1962.

4. Would have gradually increased its influence over the next 15 years because the decline continued, subtly but continuously from 1963 to the present.

5. Would not have affected achievement test takers in grades 3-11 during the recent past.

Disease vs Environment

If the SAT decline were health related, it would virtually have to be environmental (ie., social or physical or both) leading to widespread, progressive, but mild deterioration in measured cognition.

Biological factors tend to be expressed as clinical entities in varying degrees of intensity within population groups. At this time, there are no established diseases that could have had this widespread subtle effect without also producing a recognizable gradual increase in serious cases. Such disease has not been described in the literature. It is concluded, therefore, that it would not be likely for a disease process to have caused a progressive change like the SAT decline.

We have reviewed environmental factors with particular references to their potential impact on brain development during the last trimester of pregnancy and through the first 18 months of life. We have extended the inquiry through childhood as well.

With respect to the question about induced labor in pregnancy we conclude that it is unlikely to have caused the decline. Despite induction's known hazards, the decline in SAT scores does not reflect a practice that was limited to a small proportion of babies delivered each year.

The inconsistency in results from achievement tests for grades 3-11 in comparison with SAT scores also makes it extremely unlikely that there was a biological basis for the SAT decline. Any biological process capable of producing the latter would also be reflected in earlier intellectual impairment.

It is concluded from this analysis that there were no known biological or environmental factors that met the criteria established earlier in this report. Thus, the decline in SAT scores is not likely to result from physical environmental factors.

Alternate explanation. From an epidemiological perspective, the most likely alternative explanation to a disease-oriented or an environmental hypothesis is a change in the test taking population itself. If, for example, the level of education of the parents of the test-taking population were to have declined parallel to the SAT score changes, the SAT decline would not be surprising.

TECHNICAL APPENDIX

This section includes five parts; each provides a brief summary of the medical conditions, the environmental factors, and the demographic factors that were reviewed in the preparation of this report.

A bibliography is attached.

Genetic factors. Genetic disturbances can jeopardize life or cause metabolic, structural, and intellectual deficiencies. Unfortunately, in the United States, as well as in other countries, birth defects have not been consistently recorded on birth certificates. Moreover, the incidence of genetic diseases in infants' families have rarely been linked to birth outcomes. In general, the incidence of congenital defects in the United States has been declining. Infant deaths resulting from defects decreased between 1954 and 1963. In New York State, for example, in the years 1950 through 1960, the number of malformations dropped from 14.28 to 11.44/1,000 live births. The three major central nervous system anomalies (anencephalus, spina bifida, and hydrocephalus) have shown the greatest reduction. Mongolism (Down's Syndrome) experienced a slight increase in New York State (31 to 48/1,000,000 births, 1950-58).[1] However, in 1959 the extra chromosome associated with Down's Syndrome[2] was identified. Its incidence has since decreased, perhaps because of the application of genetic screening procedures. Good data could not be found for any changing incidence of diseases such as maternal diabetes or sickle cell anemia for 1945-63. Death rates because of diabetes (an important disease of pregnancy)[4,5] however, have been dropping since 1948.

Prenatal factors. Most health indicators for the prenatal period improved during 1945-1963. Complications of pregnancy and childbirth that can jeopardize the fetus accounted for fewer maternal deaths in 1963 than in 1945 (35.8 vs. 186/100,000 live births). Those attributed to sepsis, toxemia, ectopic pregnancy, and spontaneous abortion also decreased. Maternal deaths from hemorrhages rose somewhat (3.0 to 6/100,000 live births).[1] Fetal deaths from diseases and conditions of pregnancy dropped in New York State, 1953-60 (115 to 86). Deaths related to toxemia of pregnancy fell from 143 to 20 but those from other toxemias rose from 17 to 69. (This may be a result of changing classifications.)

The optimum demographic profile of a potential mother is one of a woman aged 20-34, who is married, high-school educated, over 5'4" tall, who has been non-pregnant for two years.[1] During the period 1945-60, there has been a steady decline in births to women over 40, and an increase in the percent of live births occurring in women aged 20-34.[4] The percent of illegitimate births has increased from 4.1 to 7.7% (mostly because a greater proportion of women are unmarried). The median number of school years completed by American women rose from 9.0 to 11.8 (1940-60).[16] The height of American women aged 15-44 appears to have increased by almost an inch (brining the mean height to 63.8"). Birth intervals narrowed to near two years in the 1950s, but have increased since then to almost three years. Family planning has been widely accepted since 1960, which allows more birth spacing. Although lower IQ has been shown to be associated with family size, its consideration is beyond the scope of this report.

Maternal malnutrition during pregnancy can impair fetal growth. During this period, the American diet included increased amounts of protein and carbohydrates but the sources changed. Milk, whole grains, and fresh fruits and vegetables declined in per capita consumption. There is no reason to believe that this shift in food sources has had any detrimental health consequences -- in fact, most studies suggest the nutritional status of children is improving. However, the practice of restricting weight gain during pregnancy was prevalent during 1950-70, and this may have contributed to problems associated with low birth weight infants.[10] Although low birth weight babies have lower IQ test performance as late as ages 8-10,[2,3] it does not seem sufficiently prevalent to link it to the SAT test decline.

Exposure to toxic substances are recognizable risks to fetal development. The practice of taking drugs during pregnancy was relatively common before 1963. A 1960 study by the Kaiser Permanente found 90 percent of pregnant women had taken a prescription drug.[12] Not until the thalidomide tragedy of 1963 did this matter become a public concern.[6] Another study showed that 27 percent of all women having live births in 1963 had an X-ray during pregnancy -- the majority for nonpregnancy-related reasons.[9] Smoking mothers tend to have low birth weight babies more frequently.[15] From 1955 to 1966 the proportion of women aged 18 to 44 smoking more than one package of cigarettes a day more than doubled. In addition the general incidence of women smoking in the child bearing years increased from roughly 35 to 40 percent.[15] Streissguth found that children of alcoholic mothers had more borderline to moderately retarded children by age 7 than controls (75 percent vs. 22 percent) regardless of education and even when raised in excellent foster homes.[13]

In the case of these toxic substances, however, the possible subtle forms of brain injury are not known, if they do actually exist. It seems highly unlikely that individually or collectively they could account for the pattern of change observed in the SAT test scores, and not affect achievement test scores for grades 3-11.

Labor and delivery. During the period 1945-70, many obstetrical practices, since changed, were in effect. In 1967, at least one third of all deliveries were under general anesthesia, and this frequently produced overnarcosis of the newborn. Nearly nine percent of all deliveries were electively induced. In the 1960s use of forceps peaked. There has been marked improvement in the prevention and management of Rh-factor babies.[2]

Neonatal mortality rates declined between 1950 and 1960. Those attributed to birth injuries dropped from 3.2 to 2.4/1,000 live births. The decline was primarily among normal weight infants. However, neonatal deaths from postnatal asphyxia and atelectasis increased from 3.6 to 4.5/1,000 live births.[4] However, all of the increase occurred in low birth weight babies. Since 1950 plural births (associated with more traumatic deliveries) have decreased, primarily because of the younger age distribution of mothers.[5] (Plural births tend to occur to older women.)

Declining death rates do not necessarily indicate improvement in health status. As Hellman stated: "If many babies actually die from hypoxia or trauma, during birth, it seems likely that many others escape death, but suffer permanent effects of cerebral injury. The late effects of birth injuries, both hypoxic and traumatic, are among the most important problems that require urgent and intensive study in modern obstetrics.[1,3] We do not, however, have reliable estimates of the numbers of infants so affected.

The number of in-hospital and physician-attended births increased to its present level between 1950 and 70.[6] Thus, any consequences or disadvantages of modern medical care should now be fully apparent. Any long term health implications from these changed practices have been primarily in more intensive care for high risk infants, whose survival rates did not change appreciably between 1950 and 1965.

Infancy. Low birth weight has been associated with two to three times the risk of physical defects and higher mortality. Three longitudinal studies showed at ages four, six to nine, and eight to ten that IQ was directly proportionate to birth weight regardless of social class. Among blacks, the percentages of low birth weight babies increased in the United States from 1950 to 1967. Median birth weights of whites have been stable.[6] Low birth weight infants born between 1961 and 1968 may, however, be faring better than those born earlier. They have experienced less mutual retardation, blindness, and deafness.[3]

The brain grows most perinatally and in the first two years of life. In the years 1945-65, breast feeding declined in popularity.[5] Evaporated milk was the most commonly used infant food until 1963 when the use of commercially prepared formulas became widespread.[7] Evaporated milk is higher in protein, carbohydrate, calcium, phosphorous, and vitamins D, B1 and B2 than is human milk, and lower in vitamins A, E, C, and B6, as well as iron and zinc. Also, evaporated milk must be diluted to the protein level of human milk to be digested by infants.[4] It is unlikely, however, that these changes were important neurologically, because of compensatory sources of nutrition available to American infants.

Childhood. The health of American children (ages 1-14) improved steadily between 1945 and 1960. Childhood mortality (ages 1-14) decreased between 1945 and 1955 (127.1 to 69.6/100,000 children). Deaths attributed to meningitis, influenza, gastrointestinal problems, heart disease, rheumatic fever, and accidents declined most strikingly. An increasing number of deaths were attributed to congenital malformations (4.9 to 5.8/100,000), leukemia (3.2 to 3.7), and other cancers (4.9 to 5.8).

Some other conditions seemed to increase in number, such as the measured prevalence of visual, hearing, and orthopedic impairments, perhaps because of improvements in screening. Speech defects, however, which have often been associated with mental development, had a declining prevalence in school children.[1]

Heights and weights of children increased through 1955-56 and have remained stable,[3] most likely attributable to better nutrition.

REFERENCES

Genetic Factors

1. Gittlesohn, A.M., and Milham, S. Jr., "Vital Record Incidence of Congenital Malformations in New York State," in *Genetics and Epidemiology of Chronic Disease*. Washington, D.C.: U.S. Public Health Service, 1965, pp. 305-320.

2. Lejeune, O.J., Gauthier, M., and Turpin, R., "Human Chromosomes in Tissue Culture." *Comptes Rendus Hebdomadaires des Seances de l'academie des Sciences*, Vol. 248, 1959, p. 602.

3. National Center for Health Statistics, "Infant, Fetal and Maternal Mortality, United States-1963." Vital and Health Statistics, PHS Pub. No. 1000-Series 20-No. 3, September 1966, p. 30.

4. U.S. Bureau of the Census, Historical Statistics of the United States, Colonial Times to 1957. Washington, D.C.: 1960, p. 26.

5. U.S. Bureau of the Census, Historical Statistics of the United States, Continuation to 1962 and Revisions. Washington, D.C.: 1960.

Prenatal Factors

1. Aladjem, Silvio (ed.), Risks in the Practice of Modern Obstetrics. St. Louis: C.V. Mosby Co., 1972, 304 pp.

2. Birch, H.G., and Gussow, J.D., Disadvantaged Children, Health, Nutrition and School Failure. N.Y.: Harcourt, Brace and World, Inc., 1970, pp. 60-61.

3. Francis-Williams, J., and Davies, P.A., "Low Birth Weight and IQ." Developments in Medical and Child Neurology, Vol. 16, No. 6, December 1974, pp. 709-728.

4. National Center for Health Statistics, Facts of Life and Death. Revised edition. Rockville, Md.: Public Health Service, 1970.

5. National Center for Health Statistics, "Height and Weight of Adults 18-74 years of Age in the United States." Advance Data, Vol. 3, No. 8, November 19, 1976.

6. National Center for Health Statistics, "Trends in Prematurity, United States: 1950-67." Vital and Health Statistics, PHS Pub. No. 1000-Series 3-No. 15, January 1972.

7. National Center for Health Statistics, "Changes in Cigarette Smoking Habits Between 1955 and 1966." Vital and Health Statistics, PHS Pub. No. 1000-Series 10-No. 59:8, April 1970.

8. National Center for Health Statistics, "Infant, Fetal, and Maternal Mortality, United States-1963." Vital and Health Statistics, PHS Pub. No. 1000-Series 20-No. 3:58, September 1966.

9. National Center for Health Statistics, "Medical X-ray Visits and Examinations During Pregnancy, United States-1963." Vital and Health Statistics, PHS Pub. No. 1000-Series 22-No. 5:8, June 1968.

10. National Research Council, Maternal Nutrition and the Course of Pregnancy. Washington, D.C.: National Academy of Sciences, 1970, pp. 82, 14.

11. New York (State) Department of Health, 81st Annual Statistical Report of the Department of Health for the Year Ending December 31, 1960, p. xxiv.

12. Peckham, D.H., and King, R.W., "Study of Intercurrent Conditions Observed During Pregnancy." American Journal of Obstetrics and Gynecology, Vol. 87, pp. 609-620.

13. Streissguth, A.P., "Psychologic Handicaps in Children with the Fetal Alcohol Syndrome." Work in Progress on Alcoholism, Annals of the N.Y. Academy of Sciences, Vol. 273, 1976, pp. 140-145.

14. U.S. Department of Agriculture, Nutritional Review. CFF (adm-299-9) January 1975, p. 11.

15. U.S. Department of Health, Education, and Welfare, The Health Consequences of Smoking, A Report to the Surgeon General 1971. Washington, D.C.: U.S. Government Printing Office, pp. 389-396.

16. U.S. Department of Labor, Handbook of Labor Statistics. Washington, D.C.: 1974, p. 36.

17. White House Conference on Children, Profiles of Children. Washington, D.C.: 1970, p. 96.

Labor and Delivery

1. American College of Obstetricians and Gynecologists, Survey of Obstetric Practice and Associated Services in Hospitals in the United States. Report of Committee on Maternal Health, Chicago, 1970.

2. Broman, S.H., Nicholos, P.L., and Kennedy, W.A., Preschool IQ - Prenatal and Early Development Correlates. Hillsdale, N.J.: Laurence Erlbaum Associates, 1975, p. 7.

3. Hellman, L., and Pritchard, J.A., Williams Obstetrics, 14th Edition. N.Y.: Appleton-Century Crofts, 1971, pp. 1051-20, 1138, 1009.

4. National Center for Health Statistics, "A Study of Infant Mortality from Linked Records: Comparison of Neonatal Mortality from Two Cohort Studies, United States." Vital and Health Statistics, PHS Pub. No. 1000-Series-20 No. 13, October 1967, pp. 2-3.

5. National Center for Health Statistics, "Multiple Births, United States-1964." Vital and Health Statistics, PHS Pub. No. 1000-Series 21-No. 14, October 1967, pp. 2-3.

6. National Center for Health Statistics, "Natality Statistics Analysis, United States, 1965-67." Vital and Health Statistics, PHS Pub. No. 1000-Series 21-No. 19:22, May 1970.

Infancy

1. Birch, H.F., and Gussow, J.D., Disadvantaged Children, Health, Nutrition, and School Failure. N.Y.: Harcourt, Brace and World, Inc., 1970, pp. 60-61.

2. Broman, S.H., Nichols, P.L., and Kennedy, W.A., Preschool IQ - Prenatal and Early Development Correlates. Hillside, N.J.: Laurence Erlbaum Associates, 1975, p. 3.

3. Davies, P.A., and Tizard, J.P., "Very Low Birthweight and Subsequent Neurological Defect." Developments in Medical and Child Neurology, Vol. 17 (1), February 1975, pp. 3-17.

4. J.B. Roerig Division, Handbook of Infant Formulas, 6th edition, revised. N.Y.: Charles Pfizer and Co., Inc., February 1969, 41 pp.

5. Meyer, H.G., "Breast Feeding in the United States." Clinical Pediatrics, Vol. 7, December 1968, pp. 708-715.

6. National Center for Health Statistics, "Trends in Prematurity, United States 1950-67." Vital and Health Statistics, PHS Pub. No. 1000-Series 3-No. 15, January 1972.

7. Sehring, D.S., "Infant Feeding Trends in an Industrialized Culture," in Milk and Lactation - Modern Problems in Pediatrics, Vol. 15, pp. 231-237.

Childhood

1. National Center for Health Statistics, "Children and Youth, Selected Health Characteristics, U.S. - 1958-1968." Vital and Health Statistics, PHS Pub. No. 1000-Series 10-No. 62, February 1971, p. 15.
2. U.S. Children's Bureau, "Perinatal, Infant, Childhood, and Maternal Mortality, 1955." U.S. DHEW Statistical Series, Vol. 50, 1958, p. 33.
3. U.S. Public Health Service, "New Growth Charts for American Youngsters Released by NCHS." Public Health Reports, Vol. 91, November-December 1976, p. 571.

Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates between 1960 and 1972

Albert E. Beaton
Thomas L. Hilton
William B. Schrader
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

June 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $6 per copy, as well as a price list for all the appendixes can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by Educational Testing Service.

Printed in the United States of America.

CONTENTS

Acknowledgments. 3
Overview of Findings 3
Plan of the Study. 5
    Data Sources. 5
    Variables Used. 7
    Analysis Plan 9
Changes in Reading Comprehension Scores from 1960 to 1972. 10
    Introduction. 10
    Reading Ability of High School Seniors. 11
    Reading Ability of College Entrants 16
    Reading Ability of SAT Candidates 17
    SAT Scale Drift 19
    Implications for the SAT-Verbal Score Decline 20
Changes in Other Student Characteristics from 1960 to 1972 22
    Changes in the High School Senior Population. 22
    Changes in the College Entrant Population 33
    Changes in the SAT Candidate Group. 44
Summary. 60
References 60

Appendixes:

    Appendix A. Equating the Project TALENT and National Longitudinal Study Reading Tests 63
    Appendix B. Definition of College Attendance. 77
    Appendix C. Identification of Inconsistent Data Cases 83
    Appendix D. Description of Partitioning Analysis. 85
    Appendix E. Estimated Effect of School Retention on Test Scores. 92

TABLES

1. Reading Score Levels of High School Seniors. 13
2. Reading Score Levels of College Entrants 17
3. Reading Score Levels of Students Who Took the SAT. 18
4. Reading Scores of High School Seniors Grouped by Age 23
5. Reading Scores of High School Seniors Grouped by Sex 24
6. Reading Scores of High School Seniors Grouped by Father's Education 25
7. Reading Scores of High School Seniors Grouped by Mother's Education 26
8. Reading Scores of High School Seniors Grouped by Father's Occupation. 27
9. Reading Scores of High School Seniors Grouped by Mother's Occupation. 28
10. Reading Scores of High School Seniors Grouped by Family Configuration 28
11. Reading Scores of High School Seniors Grouped by High School Curriculum 32
12. Reading Scores of High School Seniors Grouped by College Major Choice 33

13. College Entrants Grouped by Age Level 34
14. College Entrants Grouped by Sex 35
15. College Entrants Grouped by Father's Education 36
16. College Entrants Grouped by Mother's Education 37
17. College Entrants Grouped by Father's Occupation 38
18. College Entrants Grouped by Mother's Occupation 39
19. College Entrants Grouped by Family Configuration 40
20. College Entrants Grouped by High School Curriculum 42
21. College Entrants Grouped by College Major Choice 43
22. Students Who Took the SAT Grouped by Age 45
23. Students Who Took the SAT Grouped by Sex 47
24. Students Who Took the SAT Grouped by Father's Education 48
25. Students Who Took the SAT Grouped by Mother's Education 49
26. Students Who Took the SAT Grouped by Father's Occupation 51
27. Students Who Took the SAT Grouped by Mother's Occupation 52
28. Students Who Took the SAT Grouped by Family Configuration 53
29. Students Who Took the SAT Grouped by High School Curriculum . . . 56
30. Students Who Took the SAT Grouped by College Major Choice 58
31. Students Who Took the SAT Grouped by College Attendance 59

FIGURES

1. Mean Reading Score for High School Seniors, College Entrants, and SAT Takers in 1960 and 1972 12
2. Number of Students Whose Reading Score Would Place Them in Each Fifth of 1972 High School Seniors 14

ACKNOWLEDGMENTS

We wish to thank John C. Flanagan, Chairman of the Board of the American Institutes for Research, for his generous permission to use the Project TALENT data; Donald H. McLaughlin, Director, Project TALENT Data Bank, for his many hours of help and cooperation; and Marion F. Shaycoft, Research Scientist, for helpful advice. Most of all we thank Mr. Flanagan and his colleagues during the early years of Project TALENT for their foresight and expertise in conceiving and executing an excellent study of American high school students in 1960. These data along with data from the National Longitudinal Study were indispensable for the present study.

We would like to thank Senior Research Data Analyst Martha Stocking and Research Data Analysts Judith Pollack-Ohls and David H. Saxe for their extraordinarily able help and many substantive suggestions. In particular, we want to pay special tribute to Ms. Pollack-Ohls for her ability to function at peak efficiency under incredible deadlines. We would also like to thank Senior Research Data Analyst John J. Ferris for developing the procedures for maintaining subject anonymity.

Jerome A. Weisbrodt, Director, Elementary and Secondary School Programs--Program Administration, and members of his staff managed the special test administrations efficiently under time pressure.

We would like to thank our ad hoc advisors William H. Angoff, Fredric M. Lord, Gary L. Marco, and E. Elizabeth Stewart for their gracious gifts of ideas and support when needed, and to thank Mr. Angoff, Norman Frederiksen, and Ms. Stewart for their helpful reviews of the report.

We wish to thank our secretaries Susan DeMarco, Frances Livingston, Ruth Myernick, and Diane Rein for their exceptional work during a period of tight deadlines.

Support for this research was provided by Educational Testing Service and the College Entrance Examination Board. A grant from the National Institute of Education supported a feasibility study that contributed substantially to the design of the present investigation.

OVERVIEW OF FINDINGS

That the average scores of students who took the Scholastic Aptitude Test (SAT) have been declining steadily and noticeably since 1963-64 is a fact. Because the SAT candidate group is self-selected, however, there is no certainty that a parallel decline in ability occurred for all college entrants or for all high school seniors. This study was initiated in order to obtain directly comparable data on changes between 1960 and 1972 in reading ability and other characteristics for high school seniors, college entrants, and SAT takers. Similarities and differences in the changes for these groups should be useful in evaluating various explanations for the SAT score decline.

The choice of the years 1960 and 1972 for this study was determined by the fact that in each of these years there were extensive data for national probability samples of high school seniors, including follow-up data on college entrance in the year after high school graduation. For 1960 the data had been collected as part of the massive Project TALENT research effort, and for 1972 the data had been collected by the National Longitudinal Study of the High School Class of 1972 (NLS).

In each of the two studies, reasonably similar tests of reading comprehension had been given to the students. In order to use the results on these tests in comparing performance in the two years, it was necessary to equate the scores on the tests. Data for this equating were obtained by giving both tests to about 1,700 seniors in 88 high schools in 1976. We also searched Educational Testing Service (ETS) files for the SAT scores of more than 20,000 Project TALENT students. Comparable data from the two studies were available for age, sex, father's occupation, mother's occupation, father's education, mother's education, family configuration, high school curriculum, expected college major field, and college attendance.

Our research questions and findings were these:

1. Did the high school senior population change from 1960 to 1972?

The average reading score declined. Comparisons of subgroups from the 1970 and 1972 samples showed that the average 1972 senior was slightly older, more likely to have parents who completed high school and were in professional, managerial, or white collar positions, and less likely to be an only child or a firstborn child. The proportion of boys and girls became more nearly equal. In nearly every subgroup studied, there was a decline in reading score.

From 1960 to 1972, an important phenomenon in American education was a marked increase in the number of students who stayed in school instead of dropping out. A reasonable statistical estimate suggests that this trend made a considerable contribution to the observed decline in reading scores for high school seniors. Less able students who would have been dropouts in the 1960 cohort appeared as seniors in the 1972 cohort.

2. Has the college entrant population changed?

The decline in reading scores is about the same for college entrants as for high school seniors. In both years, 1960 and 1972, college entrants were substantially higher in reading ability than high school seniors. At all ability levels, the percentage of seniors going on to college increased. The 1972 entrants were slightly older, included an increased percentage of women, and were more likely to have parents who attended college and who were engaged in professional or managerial occupations than the 1960 entrants. The variation in the amount of score decline from one subgroup to another may throw some light on the sources of the overall decline, but none of the variables emerged as a major determiner.

3. Has the population taking the SAT changed?

The drop in the average reading score for SAT takers was about twice as large as the drop for college entrants and high school seniors. The percentage of students taking the SAT increased from 1960 to 1972 at all ability levels. In 1972 the SAT population had an increased proportion of women, a decreased proportion of college preparatory students, and more students from large families. As with the college entrant population, we found some variation among subgroups in the amount of decline. Of the characteristics studied, however, only the substantial decrease in the proportion of SAT takers who entered four-year colleges seems to have had an appreciable impact on the SAT score decline.

4. Has the SAT scale shifted?

Our analysis suggests that the SAT-verbal scale shifted somewhat from 1960 to 1972. Consequently, for a given level of ability a student would be expected to earn a somewhat higher SAT-verbal score in 1972 than in 1960. This result suggests that the actual decline in ability was greater than the observed decline in scores would indicate.

The SAT-verbal means for our samples are 474 in 1960 and 453 in 1972, a difference of 21 points. These means were affected by changes in the high school population, changes in the percentage of high school seniors at various ability levels who chose to take the SAT, and some drift in the scale of the SAT-verbal test. We estimate that the change in the high school population by itself would have reduced the 1960 mean by 12 points to 462, whereas the change in patterns of self-selection by candidates alone would have reduced the mean by 20 points to 454. Thus, the effect of changes in the high school population and SAT-taking patterns together resulted in a decline of 32 points to 442. The effect of scale drift was to add 11 points, resulting in the 1972 average SAT-verbal score of 453.

The largest single component of the decline in average SAT-verbal score was, therefore, changes in the percentage of high school seniors at various ability levels who chose to take the SAT. These changes are reasonable in view of the changing composition of the College Board between 1960 and 1972. Among the various student characteristics that we were able to study, several appeared to be related to the score decline for high school seniors, but much of the change in the distribution of reading ability remains to be explained.

PLAN OF THE STUDY

This study investigated differences between the 1960 and 1972 high school graduating classes--differences that may help to account for the SAT score decline. This report describes certain characteristics of three groups of high school students: all high school seniors, college entrants, and SAT candidates. The characteristics include: (1) age, (2) sex, (3) father's occupation, (4) mother's occupation, (5) father's education, (6) mother's education, (7) family configuration, (8) high school curriculum, and (9) expected college major field. The report includes the reading comprehension scores of high school seniors categorized according to the characteristics listed above, and the SAT scores of members of these subgroups for whom SAT scores were available.

Data Sources

Introduction. This study could be done because extensive data on national probability samples of United States high school seniors had been collected by Project TALENT in 1960 and by the Base-Year Survey of the National Longitudinal Study in 1972. In both studies, a follow-up provided data on college attendance. Data on SAT scores were collected as part of the NLS.

The American Institutes of Research graciously cooperated in working out an arrangement by which SAT score files could be searched for a sample of 20,359 participants in the TALENT survey while the confidentiality of both data files was maintained. The data bases for the two national surveys are well-known and well-documented. A detailed comparison of the two surveys has been made by Schrader and Hilton (1975); therefore, many details of the samples and variables will not be repeated here.

TALENT and the NLS were designed a dozen years apart for quite different purposes, at times of vastly different social emphases. TALENT was conceived in the post-Sputnik era when the nation was greatly concerned with locating talented students and encouraging their development. The National Longitudinal Study was designed to trace the careers of American students, with particular emphasis on equality of opportunity for various minority groups. The questions and the testing batteries administered to the two cohorts differed according to the needs of the study, the ingenuity of the investigators, and judgments regarding the best use of students' time. A formal comparison of the two data bases was never intended. However, when Schrader and Hilton studied these two data sets, they concluded that the reading comprehension tests of each study and a small subset of questionnaire items were sufficiently similar to permit comparisons.

The present study attempted to establish comparable scores on the reading tests used in Project TALENT and NLS and to evaluate the tests' equivalence. Because the test questions were asked and responded to at different times and their meanings to the respondents may have changed, strict comparability cannot be assumed.

Project TALENT Data Base. The Project TALENT data were collected by the American Institutes for Research in the late spring of 1960 (Flanagan et al., 1962). The project staff selected a 5 percent sample of all high school students in American public, private, and parochial schools and administered a battery of tests and a questionnaire to them. The total sample was about 400,000. The 81,175 high school seniors in the probability sample of high school seniors were sent a questionnaire during the year following graduation asking about a number of things, including what the students were doing that year. A random sample of the nonrespondents was followed by a field survey. The survey results were weighted in such a way as to make estimates of follow-up information possible for the total sample contacted originally in 1960.

A critical variable in the present study is whether the student took the College Board examinations, a variable not included in the TALENT data but available on microfilm in the ETS files. Because looking up SAT scores for the entire 81,175 sample members would have been time-consuming and expensive, and because such a large sample is statistically unnecessary, we asked the American Institutes for Research for a subsample of about 25 percent of the cases (N = 20,359), a subsample comparable in size to the entire NLS sample. The final edited data base, therefore, contains a large number of test scores, questionnaire items, and SAT scores. The data file has no individual identification information. The Project TALENT sample used in this study was selected in such a way as to be self-weighting for high school seniors. Special weights, supplied by the Project TALENT staff, were used for persons included in the sample of nonrespondents selected for special follow-up.

National Longitudinal Study Data Base. Under the aegis of the National Center for Education Statistics, the NLS staff selected a stratified random sample of American public, private, and parochial schools and in each school administered a battery of tests and questionnaires to 18 randomly selected seniors. The present study used the records of 16,683 students. SAT scores were recorded by the schools. A follow-up survey was administered two years later to learn, among other things, about the career choices of these high school graduates. About 95 percent of the sample was located. The data file, including the results of the follow-up, was obtained from the National Center for Education Statistics.

Variables Used

Reading Comprehension. Both Project TALENT and the National Longitudinal Study administered reading comprehension tests to nearly all members of the samples. Project TALENT used a test with 48 items that was scored on a number-right basis. The NLS administered a 20-item test that was scored using formula scores (number right minus 1/4 wrong). Because a comparable general test of intellectual ability is critical to the analyses discussed later, a substudy set about equating these two reading comprehension tests. Usable data were obtained for 1,657 high school seniors enrolled in 88 high schools. A detailed description and discussion of the equating study are presented in Appendix A.

In this study, reading scores were used not only to describe various subgroups but also for stratifying the three main groups on the basis of reading ability. Six strata were defined, using the distribution of reading scores for the 1972 cohort, so as to include the top tenth, the second tenth, and the second, third, fourth, and bottom fifth of the 1972 high school senior group. (In determining the strata, the 90th, 80th, 60th, 40th, and 20th percentiles for high school seniors in 1972 were calculated, expressed in terms of the NLS raw score scale. Then the corresponding TALENT raw scores were determined, using the conversion table developed in the equating study. The resulting values defined the equivalent class intervals for NLS scores and TALENT scores. In using these class intervals, frequencies were prorated between class intervals whenever an interval boundary fell within a particular NLS or TALENT raw score interval. This procedure was used in order to reduce the effect of coarseness of grouping in the raw score data.)

College Entrance. College entrance, necessarily derived from questions on the follow-up questionnaires, was defined as full-time matriculation in a two- or four-year college in the academic year following the senior year of high school. Part-time attendance or attendance in vocational or technical schools was not considered college attendance. By this definition, 39 percent of the high school seniors entered college in 1960 and 43 percent in 1972. Details on the questionnaire items used and their response frequencies are shown in Appendix B.

SAT Taking. As mentioned earlier, the SAT scores for the Project TALENT subjects had to be located in the ETS microfilm files. A search was made to locate these scores, if possible, in the ETS files for the

1958-59 or 1959-60 testing years. Of the sample, 18 percent were found to have taken the SAT at least once. Because College Board statistics were not developed for cohort groups in 1960, there is no directly comparable percentage based on all students from that cohort who took the SAT.

For the NLS sample, taking the SAT was recorded by the participating schools. The accuracy of this procedure was checked in a study by the NLS project staff (Hilton et al., 1973). In the present study, 31 percent of the sample had SAT scores, a figure very close to the 33 percent estimated from the number of persons in the class of 1972 who took the SAT, as reported by the College Board's Admissions Testing Program, divided by the number of high school seniors estimated by the National Center for Education Statistics.

SAT-Verbal and SAT-Mathematical Scores. For the Project TALENT sample, up to five sets of SAT-verbal and SAT-mathematical scores were recorded for each individual, as well as the dates of testing. The analyses used the last scores prior to high school graduation. The mean SAT-verbal score for those for whom SAT scores were available was 473.6; the SAT-mathematical mean was 496.2.

For the NLS sample, the schools were asked to record the last SAT scores before graduation. Thus, the two sets of data are comparable. For the NLS sample, the average SAT-verbal score was 453.3 and the average SAT-mathematical score 485.3, mean scores very close to the comparable figures of 453 and 484 published by the Admissions Testing Program for that year (1971-1972).

Age, Sex. In both surveys, ages were computed in months from date of birth. The average age of the 1960 sample, as of June 30 of the year of graduation, was 18 years and 1/2 month compared with an average age of 18 years and 2 months for the NLS subjects.

Sex was obtained from an item in the student questionnaire administered in each survey.

Socioeconomic Status. We used four variables from the two surveys to determine socioeconomic status and made them roughly comparable by combining categories. (These variables might be called socioeducational rather than socioeconomic.) The four variables are father's education, mother's education, father's occupation, and mother's occupation. Evidence that supports this choice of variables was recently obtained by Stricker (1976).

The four categories for education are:

1. Do not know or did not respond to the item
2. Some high school or less
3. High school graduate
4. Some college, college graduate, and graduate school

The occupational classifications are:

1. Do not know or did not respond to the item
2. *Blue collar*, which includes skilled workers or craftsmen, semiskilled workers or operatives, service workers, protective workers, laborers, all classifications of farmer, and household workers
3. *White collar*, including proprietors, salesmen, clerical workers, and military personnel
4. *Professional and managerial*, including technicians and officials

For mothers, a fifth category was included:

5. *Homemaker* or housewife

Family Configuration. Family configuration was obtained from items in each survey that, in effect, asked the respondents to indicate what their ordinal position was. Specifically, the Project TALENT Student Information Blank (SIB) asked the respondents how many living brothers, half brothers, foster brothers, or stepbrothers were older than they were and, in a second question, how many were younger. Parallel questions inquired about sisters. The respondents were instructed to include children "not now living in your home," but not to count their own twin brother or twin sister (if any).

The National Longitudinal Study included the relevant questions in the 2nd Follow-Up Questionnaire. The items asked only about brothers and sisters and gave no instructions about how to count twins, siblings not in the home, or half brothers, and so on. How the absence of these instructions affected the results, if at all, is difficult to estimate.

High School Curriculum. High school curriculum has three categories:

1. General
2. Academic or college preparatory
3. Vocational or technical

For the NLS, this information was obtained from Question 2 of the Student Questionnaire; and for Project TALENT, from Question 91 of the SIB.

Expected College Major. Creating this variable required extensive collapsing of categories. The final categories are as follows:

1. Science and mathematics, including biological sciences
2. Social sciences and humanities, including sociology, prelaw, foreign languages, and fine arts
3. Engineering
4. Other fields, including business administration, education, agriculture, nursing, home economics, and journalism
5. Nonmatching categories (categories included in one of the surveys but not the other; for example, the NLS category of Computer Science)

Analysis Plan

The analysis plan was designed primarily to facilitate key comparisons between student characteristics in 1960 and 1972 for the three main groups: high school seniors, college entrants, and SAT takers. Each of the student characteristic variables served as a basis for dividing each of the three main groups into subgroups. The particular statistics calculated for each subgroup differed for the three main groups. The following discussion of the statistical analyses has been organized on the basis of the main group or groups for which the particular calculation was made. The underlined heading for each result corresponds to the column heading for that statistic in the tables.

Statistics Calculated for All Main Groups:

"N." This is the estimated number of students in the designated subgroup and the designated main group. This result was called "Weighted N" for high school seniors, "Estimated N" for college entrants, and "N" for SAT takers. All N's are reported in thousands. Because reported percentages were calculated using N's including all digits, they will tend to differ slightly from percentages calculated from the reported N's.

"% of cohort." This result was obtained by dividing the number of students in a subgroup by the number of students in the corresponding total group. (For college entrants and SAT takers, students with missing data on a student characteristic were excluded in calculating "% of Cohort.")

"Reading mean." This is the mean reading score for members of the designated subgroup. These results make it possible to identify shifts in the composition of each of the three main groups with respect to each of the student characteristics and to determine the amount of change in reading scores from 1960 to 1972 for every subgroup.

Statistics Calculated for College Entrants and SAT Takers:

"% of stratum." This result was obtained by dividing the number of students in a subgroup (stratum) by the number of students in the corresponding subgroup of high school seniors. These results make it possible to discern trends between 1960 and 1972 in college going and in SAT-taking patterns for the subgroups defined on the basis of student characteristics. They describe changes in the patterns of *self-selection* with respect to college entrance or SAT taking. (To be sure, patterns of self-selection are heavily influenced by college policies, for example, with respect to tests required for admission.) When used along with results on "% of Cohort" for high school seniors they make it possible to assess the relative importance of changes in the high school senior population and of changes in self-selection for various subgroups in producing changes in the college entrant or SAT-taking group.

Statistics Calculated for High School Seniors Only:

"Actual N." This is simply a count of the number of sample students on which the various results for high school seniors were based.

"Reading SD." The standard deviation of reading scores was calculated to provide some information on within-subgroup variability.

Statistics Calculated for SAT Takers Only:

"SAT-verbal mean." Mean of SAT-verbal scores

"SAT-verbal SD." Standard deviation of SAT-verbal scores

"SAT-math. mean." Mean of SAT-mathematical scores

"SAT-math. SD." Standard deviation of SAT-mathematical scores

Results for "SAT-verbal mean" were of decisive importance in estimating the effect of changes in self-selection on the SAT score decline and in providing evidence on a possible shift in the SAT-verbal scale between 1960 and 1972. The other three results, although not of direct concern in this study, were considered to provide useful descriptive information on the various subgroups.

CHANGES IN READING COMPREHENSION SCORES FROM 1960 to 1972

Introduction

In this section of the report we use equated scores on the reading tests taken by Project TALENT participants in 1960 and by National Longitudinal Study participants in 1972 to explore four research questions:

1. Has the reading ability of high school seniors changed?
2. Has the reading ability of college entrants changed?
3. Has the reading ability of SAT candidates changed?
4. Does an SAT-verbal score earned in 1972 represent the same ability level as the same score earned in 1960?

We then address the question of what effect changes in the distribution of reading ability in the high school senior and SAT-taking populations have on the average score of the SAT-verbal test.

Because different reading tests were administered to the 1960 and 1972 student groups, the rigor of the comparisons reported in this chapter depends directly on the use of appropriate statistical methods for developing interchangeable scores on the two tests. Although a number of difficult technical problems were encountered in implementing the equating design,[1] we are confident that the final equating is satisfactory. We must acknowledge, however, that the equated scores do not and cannot achieve the precision that would have been attained had the same test been used in both studies.

Reading Ability of High School Seniors

Reading ability as represented by the mean scores of these two national samples of high school seniors declined between 1960 and 1972. The results may be expressed briefly by stating that the mean score declined from 10.5 to 9.7 on the NLS scale as shown in Figure 1 and Table 1. This decline amounts to about 16 percent of the standard deviation of scores for high school seniors.

Table 1 gives us the opportunity to explore the drop in average reading score further. The rows or "strata" of this table represent levels of reading ability. The six strata were defined as follows: the top tenth of the group tested in 1972 was classified in the highest group, the next tenth was classified in the second group, and the remaining fifths of the class were assigned to correspondingly lower

---

[1] One problem arose because the reading tests administered in the two studies--although fairly similar--differed in length. The TALENT test was longer than the NLS test (30 minutes vs. 15 minutes). When data collected as part of the present study were used, we found that TALENT scores were more reliable and more highly correlated with SAT-verbal scores than the NLS scores. For this reason, we decided that the initial equating of scores on the two tests was not adequate for the purposes of this study. We proceeded, therefore, to modify the TALENT scores so as to bring the correlation with SAT-verbal to approximately the same level for both tests, and to equate the modified scores. The modified TALENT scores seemed satisfactory and are used throughout this study. (The procedure used is described in more detail in Appendix A.)

Another concern was the accuracy of individual TALENT scores, particularly very low scores. On the basis of marked inconsistency of certain scores with other information in the 1960 data file, it was decided to treat about 2½ percent of the cases in the 1960 sample as having missing data on the reading test. The method by which the cases were identified is described in Appendix C.

Figure 1. Mean Reading Score for High School Seniors, College Entrants, and SAT Takers in 1960 and 1972

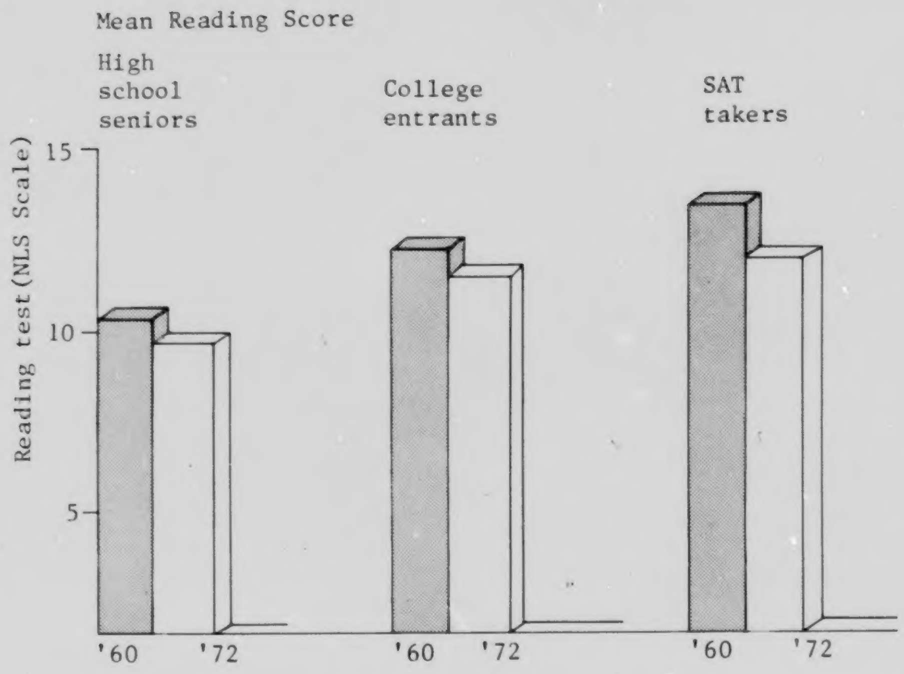


Table 1. Reading Score Levels of High School Seniors

			High School Seniors			
					Reading	
Cohort	Score	Actual N	Weighted N*	% of cohort	Mean	S.D.
TAL 1960	43.65-55	2780	255	14.3	17.8	1.1
NLS 1972	16.19-20	1442	288	10.0	17.6	1.1
TAL 1960	40.46-43.65	2356	216	12.2	15.3	0.5
NLS 1972	14.33-16.19	1504	290	10.1	15.3	0.6
TAL 1960	35.38-40.46	3905	358	20.1	12.9	0.8
NLS 1972	11.43-14.33	2956	575	20.0	12.8	0.9
TAL 1960	29.40-35.38	3683	337	19.0	10.0	0.8
NLS 1972	8.56-11.43	3042	576	20.0	10.0	1.0
TAL 1960	21.77-29.40	3554	325	18.3	6.9	1.0
NLS 1972	5.13- 8.56	3243	577	20.0	7.0	1.1
TAL 1960	(-7)-21.77	3106	284	16.0	2.4	2.4
NLS 1972	(-5)- 5.13	3649	576	20.0	2.4	2.1
TAL 1960	No test	972	89			
NLS 1972	No test	843	133			
TAL 1960	Total	20358	1864	100.0	10.5	5.1
NLS 1972	Total	16681	3015	100.0	9.7	5.0

*In thousands

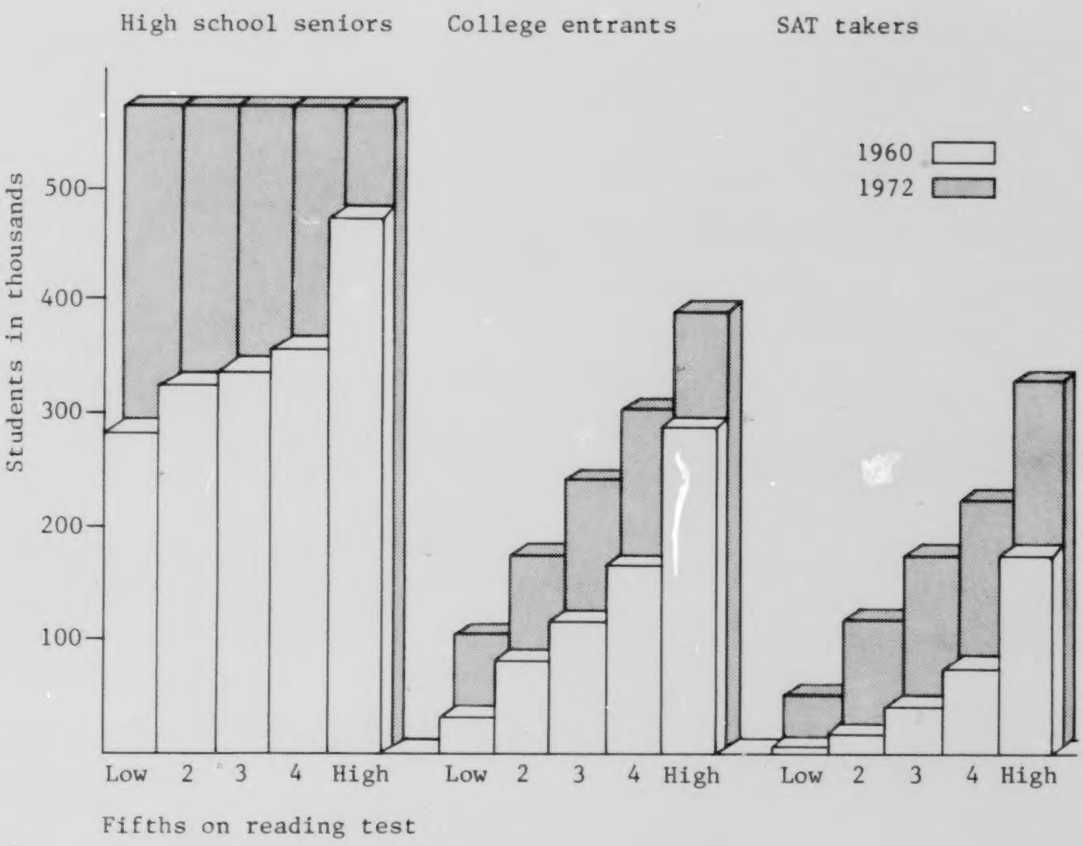
levels. On the basis of the results of the equating of 1960 and 1972 reading scores, the 1960 students were assigned to the level that best described their reading performance. Thus, each row or stratum of the table can be construed as containing students from the two samples who are substantially equal in reading ability.

The column headed "Weighted N" shows the estimated number of high school seniors in each ability stratum for the years 1960 and 1972. The section of Figure 2 headed "High School Seniors" presents the results for "Weighted N" graphically. (In the figure, the results for the two highest strata are combined.) The graph makes it evident that in 1960, the proportions of students in the higher strata are noticeably larger than the proportion of students in the lower strata.

In 1972 (as a consequence of the way in which the strata were defined), each stratum includes one-fifth of the students. Because the strata represent comparable ability levels in the two years, these results indicate that there was a decline in reading ability between 1960 and 1972 for high school seniors.

Between 1960 and 1972 the number of high school graduates increased from 1,864,000 to 3,015,000 (Simon and Frankel, 1973; Frankel and Beamer, 1974). Despite the decline in ability levels, there were actually more seniors even at the highest level, which increased by 13

Figure 2. Number of Students Whose Reading Score Would Place Them in Each Fifth of 1972 High School Seniors



percent from 255,000 to 288,000, as shown in Table 1. But the increase was much greater at the lower levels - the lowest group more than doubled, increasing from 284,000 to 576,000.

The change in the distribution is expressed in percentages in the column labeled "% of cohort." These percentages show that 14 percent of the 1960 cohort did as well or better than the top 10 percent of the 1972 cohort, and only 16 percent did as poorly as or poorer than the bottom 20 percent. This approach may give a clearer idea of the extent to which the ability level has changed than simply comparing mean scores.

The means and standard deviations of the reading scores provide evidence on the average reading ability for each stratum. The small differences between the 1960 and 1972 results indicate that the strata were, indeed, matched on ability level.

We do not know, from the data of this study, why the much larger 1972 cohort has a much larger proportion of students who scored low on the reading test. Some possible factors will be shown in the next section. We note here, however, that the change is consistent with a change in the holding power of the high schools, that is, fewer low-scoring students are dropping out of school. The percentage of the age cohort who reached the senior year of high school increased from 67 percent in 1960 to 79 percent in 1972 (Digest of Education Statistics, 1975 Edition, Table 10). Our tentative opinion is that a substantial fraction of the drop in mean reading score can be attributed to this factor, as described in Appendix E, but further research would be needed to evaluate this effect.

Other data on score changes for high school students (as distinguished from participants in national testing programs) are available from several sources. None of these sources, however, provides data for both the 1960 and the 1972 cohorts of high school graduates. The following brief discussion is concerned only with scores on reading tests and tests considered reasonably similar to reading tests, and only with scores earned by high school juniors or seniors.

Harnischfeger and Wiley (1975) present results for a number of the Iowa Tests of Educational Development for high school students in the Iowa Testing Programs. Vocabulary scores showed an increase for the 1963 through 1966 cohorts and a decline for the 1968 through 1975 cohorts. Composite scores increased for the 1963 through 1966 cohorts and declined for the 1969 through 1975 cohorts. Vocabulary scores were slightly higher, and composite scores slightly lower for the 1972 than for the 1963 cohorts.

Perry and Swanson (1974) report that scores earned by Minnesota high school juniors on the Minnesota Scholastic Aptitude Test showed an upward trend in mean score for students in the graduating classes from 1961 through 1967. A different test not equated to the earlier test was introduced for juniors in the 1968 graduating class. Students in the 1970 cohort earned mean scores somewhat higher than those for the previous two years, but for the 1971 through 1974 cohorts a perceptible and consistent score decline amounted to 14 percent of the standard deviation of the scores earned by the 1970 cohort.

Jackson and Schrader (1976) compared national Preliminary Scholastic Aptitude Test verbal norms for members of the 1962, 1968, and 1976 high school graduating classes tested in the fall of their junior year.

They found that the 1968 cohort scored noticeably higher than the 1962 and 1976 cohorts, a pattern consistent with the Iowa and Minnesota findings. If their 1976 results are adjusted for scale drift, they show a decline between the 1962 and 1976 cohorts amounting to 8 percent of the standard deviation of the 1962 cohort.

Flanagan and Jung (1971) compared the performance of national samples of members of the 1961 and 1971 graduating classes who took the TALENT Reading Test as juniors. They found a small increase in scores (about 4 percent of the standard deviation for the 1961 cohort). Flanagan (1976) also studied score changes on a number of tests between 1960 and 1975, based on a sample of 17 schools from the original TALENT sample. After a small adjustment for changes in community quality during this period, as reported by the school principals, he concluded that there had been a substantial decline on language tests, but that reading scores declined less than did vocabulary and English. Although the study report does not include data on standard deviations, it seems reasonable to assume a within-grade standard deviation of about 10 for the reading test on the basis of the 1960 Project TALENT results (Flanagan et al., 1964). The half-point decline in reading raw scores would amount to about 5 percent of a standard deviation.

Although it cannot be claimed that the foregoing summary provides a rigorous description of score trends during the 1960s and early 1970s, it suggests that scores increased during the early 1960s and that the score decline for high school seniors began a few years later than the score decline for SAT candidates. It suggests also that the score decline for high school seniors found in the present study is, if anything, somewhat greater than the decline found in other related studies.

Reading Ability of College Entrants

The reading ability of college entrants also declined between 1960 and 1972. Expressed as a mean score on the NLS scale, the decline is from 12.8 in 1960 to 11.9 in 1972. The decline in mean score for college entrants is, then, slightly more than that for high school seniors. These means can be seen in Figure 1.

This decline can be examined in some detail in the "% of cohort" column of Table 2. The percentages of college entrants in the higher strata are noticeably greater for the 1960 cohort than for the 1972 cohort. Although the percentage of college entrants in the higher strata has declined, the number (as shown in the "Estimated N" column of Table 2) in these strata has increased markedly, as it did for all high school graduates--and partly for the same reason: the large increase in the number of high school graduates. These changes are shown graphically in Figure 2. There has been a substantial increase in the number of students going to college at each level of reading ability, and these increases in the actual counts of students are more or less similar at each level. However, the net effect of this across-the-board increase in numbers of students is to make the cohort as a whole contain slightly larger percentages of low-scoring students in 1972 than in 1960, with the resulting drop in the mean. (The total in the "Estimated N" column does not equal the sum of the separate values. This discrepancy arises because the total is estimated independently of the estimates for the various strata.)

Table 2. Reading Score Levels of College Entrants

		College entrants			
Cohort	Score	Estimated N*	% of stratum	% of cohort	Reading Mean
TAL 1960	43.65-55	165	64.9	23.8	17.8
NLS 1972	16.19-20	209	72.6	16.8	17.7
TAL 1960	40.46-43.65	128	59.1	18.4	15.3
NLS 1972	14.33-16.19	187	64.7	15.1	15.3
TAL 1960	35.38-40.46	167	46.7	24.1	13.0
NLS 1972	11.43-14.33	306	53.3	24.6	12.9
TAL 1960	29.40-35.38	118	34.9	16.9	10.0
NLS 1972	8.56-11.43	246	42.7	19.8	10.1
TAL 1960	21.77-29.40	84	25.8	12.1	7.1
NLS 1972	5.13- 8.56	177	30.7	14.2	7.1
TAL 1960	(-7)-21.77	33	11.6	4.7	2.8
NLS 1972	(-5)-5.13	108	18.8	8.7	3.0
TAL 1960	No test	26	29.5		
NLS 1972	No test	56	42.0		
TAL 1960	Total	721	38.7	100.0	12.8
NLS 1972	Total	1301	43.1	100.0	11.9

*In thousands

Table 2 also shows, in the column headed "% of stratum," the percentage of high school seniors in each stratum who entered college in the year after high school graduation. Of students in the top stratum, 73 percent entered college in 1972 as compared with 65 percent in 1960. The colleges are, therefore, actually attracting a larger percentage as well as a larger number of the highest scoring students. The increase in percentages of students attending college is fairly similar for all strata.

Reading Ability of SAT Candidates

For SAT candidates the mean score on the NLS scale declined from 14.2 for the 1960 group to 12.4 for the 1972 group. This decline can be seen in Figure 1. The decline in mean score is more than twice as large as the decline for high school seniors, and twice as large as the decline for college entrants.

The column "% of cohort" of Table 3 presents the percentage of students at each ability level from 1960 to 1972. The shift is more marked for students who took the SAT than for college entrants. The percentage of SAT takers in the top stratum decreases from 34 percent

Table 3. Reading Score Levels of Students Who Took the SAT

		SAT takers							
						Sat-verbal		Sat-math	
Cohort	Score	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
TAL 1960	43.65-55	108	42.4	34.2	17.9	547	87	551	103
NLS 1972	16.19-20	177	61.4	19.5	17.7	567	84	573	97
TAL 1960	40.46-43.65	67	31.0	21.2	15.3	499	88	511	101
NLS 1972	14.33-16.19	154	53.1	16.9	15.3	504	80	527	94
TAL 1960	35.38-40.46	75	20.8	23.6	13.0	438	81	474	96
NLS 1972	11.43-14.33	226	39.3	24.9	12.9	453	78	489	96
TAL 1960	29.40-35.38	41	12.1	13.0	10.2	389	76	432	91
NLS 1972	8.56-11.43	176	30.5	19.4	10.1	404	69	446	96
TAL 1960	21.77-29.40	18	5.4	5.6	7.1	341	74	401	84
NLS 1972	5.13-8.56	121	21.0	13.4	7.1	358	68	408	93
TAL 1960	(-7)-21.77	7	2.6	2.3	2.7	337	104	385	94
NLS 1972	(-5)-5.13	54	9.4	5.9	3.2	311	61	361	86
TAL 1960	No test	11	12.7			476	114	490	113
NSL 1972	No test	27	20.7			454	115	484	105
TAL 1960	Total	326	17.5	100.0	14.2	474	108	496	110
NLS 1972	Total	935	31.0	100.0	12.4	453	108	485	113

*In thousands

in 1960 to 20 percent in 1972. In the second stratum, the drop is from 21 percent to 17 percent. The percentage in the two lowest strata (the lowest 40 percent of 1972 high school seniors) increases from about 8 percent to about 19 percent.

The "N" column in Table 3 shows the dramatic increase in the number of students taking SAT between 1960 and 1972. These figures are shown graphically in Figure 2. More students took the SAT at all levels of reading ability. In the top stratum, the number taking the SAT increased from 108,000 for the 1960 cohort to 177,000 for the 1972 cohort, while the bottom fifth increased from 7,000 to 54,000 and the next to the bottom fifth increased by 103,000. These increases at the low end of the reading ability scale result in the larger proportion of SAT takers at the lower ability levels.

The column headed "% of stratum" shows the proportion of each ability stratum taking the SAT. The proportions show increases for all strata, even the top level. In 1960, 42.4 percent of all high school seniors in the top stratum took the SAT, whereas in 1972 61.4% of that stratum did. Very few students in the lower strata took the SAT in 1960, but substantially more did in 1972. The SAT, then, is reaching more college-aspiring students in all strata. But the increase is substantially greater for students in lower ability strata.

SAT Scale Drift

Our data permitted the calculation of mean SAT-verbal and mean SAT-mathematical scores for students taking the SAT in each of the six strata. Means and standard deviations for SAT-mathematical scores are included in this report as a matter of interest. Because this study is focused on changes in verbal ability, no attempt will be made to discuss the results for SAT-mathematical. The primary objective of this phase of the analysis was to find out whether there has been a drift in the SAT-verbal score scale so that a given score represented a higher (or lower) level of ability in 1972 than in 1960.

The interpretation of the observed differences in means, however, presents a number of difficult technical problems. The questions of strict parallelism between the two reading tests and of the accuracy of some of the individual scores were noted earlier in the introduction to this chapter. The mean SAT-verbal scores for each stratum may be considered to represent a regression of SAT-verbal scores on reading scores, and Thorndike (1971) has shown that important problems arise when a regression approach is used to evaluate differences between groups. Thus, although the sample of 1,657 students in the equating study should have provided an adequate basis for equating of the kind used in this study, caution must be exercised in interpreting the differences in mean SAT scores.

A carefully designed and executed study of SAT scale drift between 1963 and 1973 has recently been completed by Modu and Stern (1976) with results relevant to the interpretation of our findings. The following table brings together the data on SAT-verbal means from Table 3 and the Modu-Stern results. The "Adjusted 1972" means apply the Modu-Stern estimates of scale drift from 1963 to 1973 to the 1972 means.

Stratum	1960	1972	Adjusted 1972
	(TALENT)	(NLS)	(Based on Modu-Stern)
Highest tenth	547	567	562
Second tenth	499	504	496
Second fifth	438	453	444
Third fifth	389	404	393
Fourth fifth	341	358	345
Lowest fifth	337	311	297
Total	474	453	444

When the SAT-verbal means for 1960 and 1972 are compared, a puzzling reversal occurs. Although the overall mean is 21 points higher for the TALENT sample than for the NLS sample, the means for all strata except the lowest show a substantial difference in the opposite direction. Within strata, the TALENT sample shows consistently lower means than the NLS sample. The column headed "Adjusted 1972," however, indicates that if the scale drift found by Modu and Stern based on the period 1963 to 1973 is used to adjust the NLS (1972) means, the results for the second through the fifth strata become satisfactorily consistent for the two samples, although the present results would suggest that the drift is if anything slightly greater for the period 1960 to 1972 than the shift found by Modu and Stern.

The difference of 15 scaled score points between the 1960 results and the adjusted 1972 results for the highest stratum remains puzzling. The standard error of the difference of the two means can be estimated, with some approximation, as 5.7 scaled score points, large enough to warrant consideration. Although imprecision of equating may have contributed to the difference, it seems probable that the main source of the difference is attributable to differences in the measurement characteristics of the two reading tests. According to this hypothesis, SAT-verbal scores for top stratum students identified by the TALENT test regressed more than did SAT-verbal scores for top stratum students identified by the NLS test.

The difference for the lowest stratum is less difficult to interpret. Multiplying the actual N (3,106) by the percentage of the cohort in the sample (2 percent) indicates that only 81 students in this stratum took the SAT. Clearly, a relatively small number of outliers in the TALENT score distribution that were not rejected by the procedure used could have raised the TALENT mean appreciably. The fact that the SAT-verbal mean score for the lowest stratum is only 4 scaled points lower than the mean for the next higher stratum is consistent with this interpretation. (The corresponding difference for the NLS sample is 47 scaled score points.)

Implications for the SAT-Verbal Score Decline

The fact that reading scores declined more than twice as much for SAT takers as for high school seniors generally makes it clear that changes in the ability level of the high school seniors can account at best for only part of the SAT-verbal score decline.

Using data already presented, we estimated the extent to which the change in the ability level of high school seniors contributed to the decline, and the extent to which the changed patterns of SAT taking contributed to it. Appendix D provides a full description of the method we used, which is called partitioning analysis. The application of partitioning analysis was based on three sets of data presented in Tables 1 and 3: the estimated number of high school seniors in each of the six ability strata in 1960 and 1972, the percentage of high school seniors in each ability stratum who took the SAT in 1960 and 1972, and the mean 1960 SAT-verbal score for students in each stratum. The results of the partitioning analysis provided answers to three questions as follows:

1. What would the SAT population be like if both the distribution of ability in the high school population and the SAT-verbal scale had remained constant, but the percentage of students at various ability levels who took the SAT changed as they in fact did between 1960 and 1972? Our estimate is that the mean verbal score of an estimated 627,000 SAT takers would be 454. The effect of the change in SAT-taking pattern alone is thus to drop the SAT-verbal mean from 474 to 454, a drop of 20 points.

2. What would the SAT population be like if the percentage of students at various ability levels who took the SAT stayed the same and the SAT-verbal scale remained constant, but the ability level of high school senior populations changed as it in fact did between 1960 and 1972? Our estimate is that the mean verbal score of an estimated 465,000 SAT takers would be 462. The effect of the change in ability level of the high school senior population alone is, therefore, to drop the SAT-verbal mean by 12 points.

3. What would the SAT population be like if the ability level of high school seniors and the percentage of students at various ability levels who took the SAT had changed as they in fact did but the SAT-verbal scale had remained constant? There would have been 935,000 SAT takers (as there were), but the mean score would have been 442 rather than 453. The net effect of scale drift is, therefore, to understate the decline in SAT-verbal mean by 11 points.

The effect of the change in the percentage of students at various ability levels who took the SAT, therefore, reduced the mean from 474, or 20 points on the SAT-verbal scale, whereas the effect of change in the ability level of high school seniors was to drop the mean from 474 to 462, by 12 points. When account is taken of scale drift, the overall decline in SAT-verbal is increased from 21 points to 32 points.

The findings of the partitioning analysis suggest that the question of why SAT scores declined can logically be thought of as two questions: Why did the decline in verbal ability of high school seniors occur? and Why did the patterns of SAT taking change? With respect to the second question, the following statement from the Report of the Commission on Tests (College Entrance Examination Board, 1970) seem remarkably pertinent:

"By 1950 there were under the relaxed [membership] requirements 115 collegiate members; by 1960 there were 350 collegiate members despite the addition of the test-use requirement in 1954; and in 1969 the collegiate membership was over 850.

"Such large increases in membership inevitably changed the character of the College Board, and it has become increasingly representative of the universe of all institutions of postsecondary education."

Further information on the nature of the changes in both the high school population and the SAT-taking population is presented in the next section, with a view to understanding the declines in mean verbal ability.

CHANGES IN OTHER STUDENT CHARACTERISTICS FROM 1960 TO 1972

In the previous section, we described and discussed changes in the measured reading ability of the high school senior population, college entrants, and students who took the SAT. In this section, nine other characteristics of these populations -- age, sex, each parent's education and occupation, family configuration, curriculum, and expected college major (if any)--will be described as a means of probing further into some changes that may help us to understand shifts in ability. In this section, all characteristics will be considered first for high school seniors, then for college entrants, and finally for SAT candidates.

Changes in the High School Senior Population

Age. As shown in Table 4, the period from 1960 to 1972 saw a marked decrease in the percentage of high school seniors in the 17½ and younger age group. This decrease may have arisen because fewer children were entering school at a relatively early age, or because there was less tendency to accelerate rapid learners. An alternative hypothesis is that the reduced percentage of younger students may have resulted, in part, from a reduced tendency for students who were progressing through school at an average or slower rate to drop out. From this viewpoint, the increased holding power of the schools during this period may have contributed to this result.

The fact that the *increase* in the percentage of high school seniors who are between 18 and 18½ is almost exactly equal to the *decrease* in the percentage who are under 17½ suggests that changed attitudes toward acceleration of school progress is the main source of the change in the percentage of younger students.

Table 4 also shows that there were small increases in the percentage of students in the two older groups, perhaps reflecting the increase in the holding power of high schools during this period.

Results for mean reading scores show that within the twelfth grade groups in both years, younger students tend to earn higher scores than older students, presumably because abler students progress more rapidly through school than do less able students. Despite the shift in the age distribution between 1960 and 1972, the score decline within each age group is similar to the overall decline.

Table 4. Reading Scores of High School Seniors Grouped by Age

	High school seniors				
Age	Actual N	Weighted N*	% of cohort	Reading	
				Mean	S.D.
17½ or less					
1960	2971	272	14.6	11.7	4.8
1972	927	166	5.5	11.0	5.0
17½+ to 18					
1960	7900	723	38.8	11.3	4.8
1972	6395	1171	38.8	10.5	4.7
18+ to 18½					
1960	6003	550	29.5	10.7	4.9
1972	6246	1162	38.5	10.1	4.8
18½+ to 19					
1960	1824	167	9.0	8.5	5.1
1972	1817	314	10.4	8.0	5.2
Above 19					
1960	1180	108	5.8	6.2	5.0
1972	1173	184	6.1	5.2	4.5
No response					
1960	481	44	2.4	7.8	5.6
1972	125	19	0.6	4.5	4.1
Total					
1960	20359	1864	100.0	10.5	5.1
1972	16683	3015	100.0	9.7	5.0

*In thousands

The role of age changes in the score decline for high school seniors is difficult to assess. The proportion of students belonging to the younger, higher-scoring strata decreased and the proportion of students belonging to the older, lower-scoring strata increased from 1960 to 1972. This result suggests that age changes contributed slightly to the score decline. To the extent that the less rapid progress of the 1972 cohort through school resulted from changed attitudes toward acceleration, however, the effect of age changes may be even smaller than the results indicate.

Sex. The percentage of young men in these samples increased by one point, from 49 percent to 50 percent (see Table 5), perhaps because of a change in the dropout rate--a change that resulted in more young men staying in high school until graduation. The greater increase in the

Table 5. Reading Scores of High School Seniors Grouped by Sex

Sex	High School Seniors Actual N	 Weighted N*	 % of cohort	Reading Mean	 S.D.
Males					
1960	9938	910	48.8	10.6	5.2
1972	8275	1502	49.8	9.6	5.0
Females					
1960	10421	954	51.2	10.5	5.0
1972	8395	1511	50.1	9.8	5.0
No response					
1960					
1972	13	2	0.1	8.4	4.5
Total					
1960	20359	1864	100.0	10.5	5.1
1972	16683	3015	100.0	9.7	5.0

*In thousands

number of men in comparison to the number of women may also have resulted in the slightly larger drop in the mean reading score for men, a drop of 1.0 compared to a drop of .7 for women. The suggestion is that in 1972 slightly more low-ability men were graduating than in 1960. However, although even a small change may be socially important, the possible contribution to the overall score decline is negligible.

Parent's Education and Occupation. A general United States population trend has been for parents to be better educated and more likely to be in professional or managerial occupations. The tables on father's and mother's education (Tables 6 and 7) show that in 1972 over 8 percent more students than in 1960 indicated that their father had some college; over 10 percent more indicated that their father graduated from high school; and nearly 14 percent fewer indicated that their father had not graduated from high school. Similarly, over 5 percent more indicated that their mother had some college; over 10 percent more indicated that their mother graduated from high school; and over 13 percent fewer indicated that their mother had not completed high school.

The same is true for almost all occupational groupings (see Tables 8 and 9). The tables show an increase in the percentage of mothers and fathers in professional and white collar jobs and a decrease in the percentage in blue collar jobs. (This comparison is confounded, however, by the fact that Project TALENT instructed students not to check "Housewife" if the mother had worked for pay during the past three years, whereas NLS permitted the student to choose the mother's principal occupation.)

With respect to reading comprehension, a finding of some interest is the relatively large drop of 1.3 in the mean reading score for the students whose fathers did not continue their education beyond high school.

Table 6. Reading Scores of High School Seniors Grouped by Father's Education

Father's Education	High school seniors				
	Actual N	Weighted N*	% of cohort	Reading Mean	Reading S.D.
Post high school					
1960	4394	402	21.6	12.6	4.7
1972	4976	896	29.7	11.9	4.6
High school graduate					
1960	4122	377	20.2	11.4	4.7
1972	4378	913	30.3	10.1	4.6
Less than high school graduate					
1960	8205	751	40.3	9.6	4.9
1972	4700	795	26.4	8.8	4.7
Don't know; blank					
1960	3638	333	17.9	9.2	5.5
1972	2629	411	13.6	5.8	4.7
Total					
1960	20359	1864	100.0	10.5	5.1
1972	16683	3015	100.0	9.7	5.0

*In thousands

Also noteworthy is the fact that the percentage of the cohort whose fathers did not complete high school decreased from 40 to 26. Because this was a low-scoring group in both years, this shift would tend to reduce the score decline.

Similarly, for father's occupation, the finding of interest is the increase in holders of higher-level positions whose children show a greater decrease in mean scores than the students with fathers in blue collar jobs.

The major shift for mothers was from blue collar to white collar positions. There was a greater decrease in the mean reading score of students with mothers in "white collar" positions than for those with mothers classified as "blue collar."

Family Configuration. In an article that has received wide attention, Zajonc (1976) proposed that intellectual development is closely related to the configuration of the family in which the individual child matures, since this configuration strongly influences the child's intellectual environment. "Intellectual environment," according to Zajonc, "can be thought of in this context as being some function of the average of the absolute intellectual levels of its members" (p. 227).

Table 7. Reading Scores of High School Seniors Grouped by Mother's Education

	High school seniors				
				Reading	
Mother's education	Actual N	Weighted N*	% of cohort	Mean	S.D.
Post high school					
1960	3995	366	19.6	12.5	4.7
1972	4006	745	24.7	11.9	4.5
High school graduate					
1960	5942	544	29.2	11.2	4.8
1972	6123	1194	39.6	10.3	4.6
Less than high school graduate					
1960	7491	686	36.8	9.4	5.0
1972	4245	715	23.7	8.6	4.8
Don't know; blank					
1960	2931	268	14.4	9.1	5.6
1972	2309	361	12.0	5.6	4.7
Total					
1960	20359	1864	100.0	10.5	5.1
1972	16683	3015	100.0	9.7	5.0

*In thousands

In general, the more siblings that children have, the lower the level of their intellectual environment and, thus, the rate of their intellectual development.

Zajonc further hypothesizes that the precise effect of the configuration depends on the time interval between successive siblings. A much older sibling can substantially raise the intellectual level of the environment. A child with a 20-year-old sibling has, in effect, another adult in his or her environment.

The effect is also mediated by the extent to which a young person is responsible for teaching his or her siblings. Zajonc hypothesizes that such teaching accelerates the intellectual development of the teacher. For this reason, only children and last-born children are hypothesized to develop at a lower rate than children in other positions.

These propositions have special relevance to national trends in test scores. Zajonc presents data showing a correspondence between mean SAT scores in the United States from 1963 to 1975 and the average order of live births 18 years earlier. Accordingly, he has proposed that "aggregate family factors are deeply implicated in the declining SAT scores" (p. 227).

Table 8. Reading Scores of High School Seniors Grouped by Father's Occupation

Father's occupation	High school seniors Actual N	Weighted N*	% of cohort	Reading Mean	S.D.
Professional, managerial					
1960	3399	311	16.7	12.8	4.6
1972	3769	732	24.3	11.9	4.6
White collar					
1960	2815	258	13.8	12.0	4.5
1972	2342	437	14.5	10.8	4.6
Blue collar					
1960	10666	977	52.4	9.8	4.9
1972	6696	1213	40.2	9.7	4.7
Don't know; blank					
1960	3479	319	17.1	9.3	5.5
1972	3876	634	21.0	6.6	4.7
Total					
1960	20359	1864	100.0	10.5	5.1
1972	16683	3015	100.0	9.7	5.0

*In thousands

The present study provided a unique opportunity to test Zajonc's model (the confluence model) for the period from 1960 to 1972 by means of the family configuration data described in the first section of this paper.

Table 10 presents the results for the high school senior population. Our first observation is that the distribution of family configurations did, indeed, change from 1960 to 1972. The percentage of only children decreased from 8 to 3 and the percentage of first children in families of two decreased from 13 to 7. For the balance of the configurations, the percentages of first-born children consistently decreased and the percentages of all later born children increased.

As for mean reading scores, first consider the Project TALENT results. The correspondence between the TALENT means and the Belmont-Marolla results reported by Zajonc and Markus (1975) is extraordinary. In every important respect the two samples agree: (1) mean scores decline with family size; (2) within each family size, the scores decline with birth order, with the exception of the next to the last child in families of four and five who exhibit a slight quadratic upswing; (3) within each family size the last born has the lowest mean; and (4) those who are

Table 9. Reading Scores of High School Seniors Grouped by Mother's Occupation

	High school seniors				
				Reading	
Mother's occupation	Actual N	Weighted N*	% of cohort	Mean	S.D.
Professional, managerial					
1960	1513	139	7.4	11.8	4.9
1972	1557	283	9.4	11.5	4.7
White collar					
1960	2611	239	12.8	12.4	4.4
1972	2672	527	17.5	11.2	4.5
Blue collar					
1960	3604	330	17.7	9.6	5.0
1972	1742	302	10.0	9.8	4.7
Homemaker					
1960	9807	898	48.2	10.5	5.0
1972	7520	1368	45.4	9.9	4.9
Don't know; blank					
1960	2824	259	13.9	9.3	5.6
1972	3192	535	17.7	6.9	4.8
Total					
1960	20359	1864	100.0	10.5	5.1
1972	16683	3015	100.0	9.7	5.0

*In thousands

Table 10. Reading Scores of High School Seniors Grouped by Family Configuration

	High school seniors				
				Reading	
Birth order and family size	Actual N	Weighted N*	% of cohort	Mean	S.D.
Only child					
1960	1520	139	7.5	11.5	4.7
1972	495	96	3.2	10.8	4.9
First of two					
1960	2574	236	12.6	12.2	4.5
1972	1193	220	7.3	11.1	4.7

(Continued)

Table 10 (Continued)

Birth order and family size	High school seniors				
	Actual N	Weighted N*	% of cohort	Reading	
				Mean	S.D.
Second of two					
1960	1694	155	8.3	11.3	4.8
1972	1356	262	8.7	10.6	4.6
First of three					
1960	1813	166	8.9	11.8	4.6
1972	1106	214	7.1	11.1	4.7
Second of three					
1960	1210	111	5.9	10.9	4.7
1972	1234	234	7.8	10.3	4.7
Third of three					
1960	837	77	4.1	10.7	4.9
1972	1127	203	6.7	9.9	4.9
First of four					
1960	1028	94	5.0	11.6	4.8
1972	700	131	4.4	10.8	4.8
Second of four					
1960	771	71	3.8	10.4	4.9
1972	800	144	4.8	9.9	4.8
Third of four					
1960	519	48	2.5	10.5	4.9
1972	795	150	5.0	10.1	4.8
Fourth of four					
1960	422	39	2.1	10.3	4.9
1972	512	96	3.2	10.1	4.9
First of five					
1960	522	48	2.6	11.0	5.1
1972	401	71	2.4	10.5	4.3
Second of five					
1960	385	35	1.9	10.1	4.9
1972	406	72	2.4	9.3	5.2
Third of five					
1960	321	29	1.6	9.9	4.7
1972	476	84	2.8	9.0	5.3

(Continued)

Table 10 (Continued)

Birth order and family size	High School seniors				
	Actual N	Weighted N*	% of cohort	Reading	
				Mean	S.D.
Fourth of five					
1960	246	23	1.2	10.2	4.6
1972	325	57	1.9	9.4	4.8
Fifth of five					
1960	241	22	1.2	9.0	5.2
1972	225	38	1.3	9.9	5.0
Six or more children					
1960	3419	313	16.8	8.8	5.1
1972	3383	566	18.8	8.6	5.1
No data					
1960	2837	260	13.9	9.0	5.8
1972	2149	376	12.5	8.0	5.1
Total,					
1960	20359	1864	100.0	10.5	5.1
1972	16683	3015	100.0	9.7	5.0

*In thousands

only children score slightly lower than the first born in families of four. In the Belmont-Marolla data, the highest and the lowest scores are separated by about two-thirds of one standard deviation. In the TALENT data, the difference between the highest mean (for the first born of two) and the lowest (the mean for all subjects in families of six or greater) is exactly two-thirds of one standard deviation. The only observable way in which the TALENT sample differs is that, unlike the Belmont-Marolla data, the decline from the next-to-last child to the last child is not in all cases greater than the average difference between other adjacent siblings.

The means for the NLS sample are similar in two respects: they decline with family size, and the mean for only children is equal to the mean for the first born in families of four. However, mean scores for the last children in families of four and five do not decline, as do the TALENT and Belmont-Marolla means. In fact, mean scores for the fourth and fifth children in families of five exhibit a marked upswing. In this respect, the NLS means are similar to the Scottish data reported by Zajonc (1976). Although a full investigation of these differences is beyond the scope of the present study, we can speculate that the differences are related to the time interval between the later born children in the NLS families of four and five.

For present purposes, the important question is how much the change in the distribution of family configurations may have contributed to the decline in mean reading scores for high school seniors. There no doubt was some effect, for in general the higher scoring family configurations declined in number relative to the lower scoring configurations. It is equally clear, however, that the effect accounts for only a fraction of the decline because all configurations except one (children who were fifth in families of five) showed declines in mean reading scores that are similar to the overall decline. In other words, when family configuration is held constant, declines in reading scores are still observed.

Because of the widespread interest in Zajonc's confluence model we attempted to estimate the magnitude of the observed effects. This was done by estimating what the 1972 mean might have been if nothing had changed from 1960 to 1972 except for the percentages of the student cohort falling in each category of family configuration (i.e., the decrease in only children from 8 percent to 3 percent, the decrease of first of two's from 13 percent to 7 percent, the increase of second of two's from 8 percent to 9 percent, and so on). This is provided by the vector product of the 1972 "% of cohort" column and the 1960 "reading mean" column. This product indicates that the 1972 overall mean reading score would have been 10.3, rather than 10.5 as it was in 1960. This shift is about 4 percent of the standard deviation of the reading scale. Since the observed 1972 mean was 9.7, or .8 of a score point less than the 1960 mean, the decline of .2 is about one-fourth as large as the observed decline in the reading scores of the high school seniors.

High School Curriculum. In 1972, somewhat more students were enrolled in academic and general courses and proportionately fewer in the vocational-technical area (see Table 11). However, the students in each category show approximately the same decline in mean scores, suggesting that curriculum is not a significant variable so far as the general decline is concerned.

Expected College Major Field. In each total sample, the number of students who expected to major in science or mathematics declined slightly (from 9 percent to 7 percent); social sciences and humanities increased considerably (from 9 percent to 15 percent); engineering declined (from 7 percent to 3 percent); and "other matching fields" remained approximately constant (see Table 12).

In general, the students who planned to go to college and who had tentatively selected a major either show no decline--the mathematics and science students actually gained in reading comprehension--or less decline than the population of seniors. The percentage of students who were not planning to attend college or were planning to do so but did not report a major field choice, increased from 1960 to 1972, and the mean score for this group dropped from 9.0 to 7.7, a decline appreciably greater than the overall decline. These results raise an important question: Why did students whose future plans showed a clear academic orientation show little or no score decline from 1960 to 1972?

Summary of Changes in High School Senior Population. The second section of this paper reported an appreciable general decline in reading comprehension. Numerous hypotheses could be advanced to explain the decline: for example, a decline in the effectiveness of instruction, the effect of television watching, student disenchantment with education, or a change in the composition of the high school graduating classes. This section has been devoted largely to an investigation of the last hypothesis.

Table 11. Reading Scores of High School Seniors Grouped by High School Curriculum

High school curriculum	High school seniors: Actual N	Weighted N*	% of cohort	Reading: Mean	S.D.
General					
1960	4194	384	20.6	9.2	4.8
1972	5665	957	31.7	8.3	4.8
Academic					
1960	7825	716	38.4	13.1	4.3
1972	6811	1379	45.7	11.9	4.5
Vocational-technical					
1960	6692	613	32.9	8.4	4.8
1972	4201	678	22.5	7.3	4.6
No response					
1960	1648	151	8.1	10.4	5.4
1972	6	1		5.3	2.3
Total					
1960	20359	1864	100.0	10.5	5.1
1972	16683	3015	100.0	9.7	5.0

*In thousands

Probably the most salient finding is the pervasiveness of the decline. The decline in reading comprehension scores was found in almost every subgroup of the high school population examined. Older students and younger students, boys and girls, students from high SES families and students from low SES families, only children and students with many siblings, college preparatory students, and noncollege students--all showed a decline. This would argue for some general influence on the attainment of all high school seniors. We did, however, observe a number of changes in the composition of senior classes from 1960 to 1972 that may have contributed to the observed decline. In summary, the relevant changes were as follows:

1. A small increase in age
2. An increase in fathers who completed high school but did not continue their education
3. A decrease in the percentage of only children and first-born children
4. An increase in the proportion of students who either did not intend to attend college or intended to attend but did not report a choice of major field

Table 12. Reading Scores of High School Seniors Grouped by College Major Choice

	High school seniors				
				Reading	
Expected college major	Actual N	Weighted N*	% of cohort	Mean	S.D.
Science, mathematics					
1960	1746	160	8.6	12.9	4.8
1972	1099	211	7.0	13.2	3.9
Social sciences, humanities					
1960	1910	175	9.4	13.1	4.6
1972	2359	453	15.0	12.4	4.3
Engineering					
1960	1498	137	7.4	12.2	4.8
1972	413	77	2.6	11.9	4.3
Other matching fields					
1960	5003	458	24.6	11.1	4.8
1972	3546	675	22.4	11.1	4.4
Nonmatching fields**					
1960	1331	122	6.5	10.2	4.9
1972	416	77	2.5	10.1	4.4
No college plans; no response					
1960	8871	812	43.6	9.0	5.0
1972	8850	1522	50.5	7.7	4.8
Total					
1960	20359	1864	100.0	10.5	5.1
1972	16683	3015	100.0	9.7	5.0

*In thousands

**Fields that were listed in one survey but had no counterpart in the other survey

Changes in the College Entrant Population

In what way has the collection of students going to college changed? Certainly, American colleges have made a concerted effort to be more egalitarian in admissions policies and have searched for and welcomed many students of lower socioeconomic background. What effect, if any, has this effort had?

The reader should be reminded that the percentages of college entrants reported here may be lower than the percentages reported elsewhere be-

Table 13. College Entrants Grouped by Age Level

	College entrants			
Age	Estimated N*	% of stratum	% of cohort	Reading mean
17½ or less				
1960	123	45.4	17.3	13.6
1972	89	53.5	6.8	12.8
17½+ to 18				
1960	315	43.6	44.2	13.2
1972	547	46.8	42.1	12.2
18+ to 18½				
1960	213	38.8	30.0	12.6
1972	533	45.9	41.0	11.9
18½+ to 19				
1960	38	22.8	5.3	11.4
1972	97	30.9	7.5	10.8
Above 19				
1960	22	20.7	3.1	9.0
1972	29	15.9	2.2	7.6
No response				
1960	9	20.0		11.9
1972	1	7.3		3.6
Total				
1960	721	38.7	100.0	12.8
1972	1301	43.1	100.0	11.9

*In thousands

cause our figures include only those who attended college *full-time in the year immediately following their graduation from college*. Looking at the bottom row of Table 13, we see that, according to our definition, the percentage of graduating high school seniors who attended college increased from 39 percent in 1960 to 43 percent in 1972.

Age. Except for the 19-years-and-older group, the percentage in each stratum going to college is higher in 1972 than in 1960. In both 1960 and 1972, the percentage of students going to college decreases as age increases. Similarly, the mean reading score decreases as age increases.

As in the high school senior group, the percentage of students 17½ and younger in the college entrant group decreased markedly, from 17 percent to 7 percent, between 1960 and 1972. Also as in the high school senior group, the percentage of students in the 18-to-18½-year group increased by about the same amount, from 30 percent to 41 percent.

The noticeable reduction in the percentage of students in the higher scoring youngest age group and the corresponding increase in the lower scoring 18-to-18½-year group may have contributed slightly to the score decline. For college entrants as for college seniors, however, the contribution of age changes to the score decline is difficult to assess because the age changes may have resulted mainly from changed attitudes toward acceleration.

Sex. Changes in the role of women in our society have resulted in a sharp change in the ratio of sexes in the college entrant group. Males made up 57 percent of the 1960 cohort but only 52 percent of the 1972 cohort (see Table 14). The composition of the college entrant population, therefore, changed substantially over these 12 years. With respect to reading tests, males scored slightly higher in 1960 and females slightly higher in 1972. The changes in the representation of the sexes in the college entrant population and the shift in relative ability support the view that the talent loss among bright girls was greater than that for boys in 1960. By 1972, the difference in college-going percentage between boys and girls has been markedly reduced, a finding of considerable importance. The observed shifts do not seem, however, to help explain the score decline.

Parent's Education and Occupation. The percentages of students entering college have not changed substantially as a function of father's education. About the same percentage of children of college-educated fathers, high-school-educated fathers, and "less-than-high-school"-

Table 14. College Entrants Grouped by Sex

	College entrants			
Sex	Estimated N*	% of stratum	% of cohort	Reading mean
Males				
1960	410	45.1	56.9	12.9
1972	675	45.0	51.9	11.7
Females				
1960	314	32.9	43.5	12.7
1972	626	41.4	48.1	12.0
No response				
1960				
1972		19.2		11.3
Total				
1960	721	38.7	100.0	12.8
1972	1301	43.2	100.0	11.9

* In thousands

Table 15. College Entrants Grouped by Father's Education

	College entrants			
Father's education	Estimated N*	% of stratum	% of cohort	Reading mean
Post high school				
1960	263	65.3	42.7	13.8
1972	575	64.2	47.7	13.0
High school graduate				
1960	165	43.8	26.9	13.2
1972	400	43.9	33.2	11.6
Less than high school graduate				
1960	196	26.1	31.8	11.8
1972	225	28.3	18.7	11.0
Don't know; blank				
1960	106	31.9		11.9
1972	96	23.4		8.1
Total				
1960	721	38.7	100.0	12.8
1972	1300	43.1	100.0	11.9

*In thousands

educated fathers go to college (see Table 15). Also, when students are grouped by mother's education, the percentage going to college is somewhat higher both for students whose mothers attended college and for those whose mothers did not complete high school (see Table 16).

The numbers of students with parents at the various educational levels changed, however, with the result of a change in the distribution of parents of students going to college, as shown in the "% of cohort" column. In 1972, nearly 48 percent of the reporting students had fathers who had at least some college, against 43 percent in 1960; the percentage with fathers who had a high school education went from 27 percent to 33 percent over the 12 years; and the percentage of college entrants who had fathers with less than a full high school education dropped from 32 percent to 19 percent. The change in the distribution of mother's education is similar. A larger percentage of college entrants now have mothers with some college education or a high school diploma, and fewer have mothers who did not graduate from high school.

College entrants whose fathers completed high school but did not attend college showed a noticeably greater score decline than the total group. A similar pattern was found for high school seniors (Table 6). On the other hand, college entrants whose mothers did not graduate from high school showed a relatively large score decline, but this subgroup

Table 16. College Entrants Grouped by Mother's Education

	College entrants			
Mother's education	Estimated N*	% of stratum	% of cohort	Reading mean
Post high school				
1960	223	60.9	34.8	13.8
1972	489	65.6	40.0	12.9
High school graduate				
1960	244	44.9	38.2	12.9
1972	517	43.3	42.3	12.0
Less than high school graduate				
1960	170	24.8	26.6	12.0
1972	210	29.4	17.2	10.6
Don't know; blank				
1960	82	30.5		11.9
1972	79	22.0		7.8
Total				
1960	721	38.7	100.0	12.8
1972	1301	43.1	100.0	11.9

*In thousands

did not show an unusually large decline in the analysis of high school seniors (Table 7). Thus, the results for mother's education suggest that factors in self-selection for college attendance may have changed for this subgroup between 1960 and 1972.

The effect of the change in parent's occupation is not so clear. Actually, a smaller percentage of the high school students who reported that their fathers were in professional or managerial positions entered college, and a slightly smaller percentage of students with fathers in white collar jobs entered college (see Table 17). A larger percentage of students with fathers in blue collar jobs entered college. For all four groups of students who reported mother's education, there was an increase in the percentage who entered college, the largest increase occurring for students whose mothers held blue collar jobs (see Table 18).

The change in the general distribution of parent's occupational status has also had an effect on the total composition of college classes. The percentage of college freshmen with fathers in professional or managerial jobs increased from 35 percent to 41 percent, and the proportion whose fathers have white or blue collar jobs decreased slightly--as did the proportion who have mothers in blue collar jobs or at home.

Table 17. College Entrants Grouped by Father's Occupation

	College entrants			
Father's occupation	Estimated N*	% of stratum	% of cohort	Reading mean
Professional, managerial				
1960	218	70.1	34.8	13.5
1972	461	63.0	40.8	13.1
White collar				
1960	139	53.9	22.2	13.5
1972	224	51.2	19.8	12.3
Blue collar				
1960	267	27.3	42.6	12.2
1972	441	36.4	39.0	11.6
Don't know; blank				
1960	95	29.7		12.1
1972	170	26.8		8.8
Total				
1960	721	38.7	100.0	12.8
1972	1301	43.1	100.0	11.9

*In thousands

College entrants whose fathers were professionals or managers showed a noticeably smaller score decline than the overall group. Because the proportion of the college entrant cohort belonging to this subgroup increased from 1960 to 1972, the trends for this group would tend to reduce the score decline. That the percentage of high school seniors in this subgroup who entered college declined from 70 percent in 1960 to 63 percent in 1972 suggests that self-selection may have contributed to this result.

With respect to mother's occupation, college entrants whose mothers were employed as blue collar workers showed a somewhat greater score decline than the total group. This presumably reflects a changed pattern of self-selection, in light of the fact that the reading score of the high school seniors in this subgroup increased slightly.

To sum up parental changes, about the same percentage of students of each parental group attended college, but the change in the distribution of parent's occupation resulted in a moderate change in the composition of the student body. The parents of college entrants would seem to be, as a whole, slightly better educated and slightly more likely to be engaged in professional and managerial occupations in 1972 than in 1960.

Family Configuration. In general, the relationship between family configuration and college entrance is what would be predicted on the

Table 18. College Entrants Grouped by Mother's Occupation

Mother's occupation	College entrants Estimated N*	% of stratum	% of cohort	Reading mean
Professional, managerial				
1960	77	55.7	12.1	13.4
1972	175	61.9	15.3	12.8
White collar				
1960	119	49.7	18.6	13.5
1972	266	50.5	23.2	12.5
Blue collar				
1960	82	24.8	12.8	12.8
1972	104	34.5	9.1	11.6
Homemaker				
1960	362	40.3	56.6	12.6
1972	596	43.6	52.1	12.1
Don't know; blank				
1960	82	31.8		12.5
1972	156	29.2		9.2
Total				
1960	721	38.7	100.0	12.8
1972	1301	43.1	100.0	11.9

*In thousands

basis of the relationship between family configuration and developed ability. The students in configurations with higher means exhibit a higher probability of entering college. As shown in Table 19, the first children in families of two, the group among the high school senior population with the highest mean, have the highest percentages attending college (52 percent for Project TALENT and 58 percent for the NLS). In both samples, the larger the family, the lower the average percentage entering college. The differences are less for NLS subjects, perhaps reflecting increased opportunity for higher education in 1972.

Within families, college entrance declines with increasing ordinal position except for an upswing in college entrance for the third and fourth child in families of four and the last child in families of five.

From 1960 to 1972 the percentage of college entrants in all configurations increased, but by different degrees. The increases range from 3 percent for the "Fourth of four" stratum to 18 percent for the "Fourth of five" stratum. No pattern in the increases emerges except, possibly, for generally larger increases for students in larger families. Again,

Table 19. College Entrants Grouped by Family Configuration

	College entrants			
	Estimated N*	% of stratum	% of cohort	Reading mean
Only child				
1960	67	48.4	10.5	13.0
1972	55	56.9	4.6	12.4
First of two				
1960	123	52.3	19.2	13.4
1972	129	58.5	10.7	12.5
Second of two				
1960	76	48.7	11.8	12.8
1972	140	53.4	11.6	11.8
First of three				
1960	79	47.8	12.4	13.6
1972	112	52.4	9.3	12.7
Second of three				
1960	50	45.5	7.9	12.7
1972	114	48.7	9.5	12.2
Third of three				
1960	30	38.9	4.6	13.2
1972	93	45.6	7.7	11.8
First of four				
1960	40	42.7	6.3	13.9
1972	61	46.5	5.1	12.5
Second of four				
1960	21	29.2	3.2	12.9
1972	61	42.6	5.1	11.4
Third of four				
1960	17	36.2	2.7	12.3
1972	68	45.3	5.7	12.3
Fourth of four				
1960	16	40.3	2.4	12.2
1972	41	43.0	3.4	12.1
First of five				
1960	14	30.0	2.2	14.2
1972	30	41.8	2.5	11.8

(Continued)

Table 19. (Continued)

	College entrants			
	Estimated N*	% of stratum	% of cohort	Reading mean
Second of five				
1960	9	25.5	1.4	14.1
1972	27	36.7	2.2	11.8
Third of five				
1960	7	25.4	1.2	12.1
1972	31	36.5	2.5	11.9
Fourth of five				
1960	4	18.7	0.7	13.1
1972	21	36.8	1.7	11.8
Fifth of five				
1960	8	36.3	1.2	12.8
1972	15	39.3	1.2	11.5
Six or more children				
1960	80	25.4	12.4	11.1
1972	192	33.8	15.9	11.0
No data				
1960	79	30.6	0.0	12.2
1972	111	29.6	0.0	10.5
Total				
1960	721	38.7	100.0	12.8
1972	1314	43.6	100.0	11.9

*In thousands

this may reflect expanded educational opportunities. The mean reading scores for each stratum of college entrants are consistent with this speculation. In the 1960 sample, the mean scores for college entrants in the larger families were generally high. It is plausible that these relatively high means were the result of more stringent selection on ability for college entrants in the larger families. But this pattern disappears in the NLS means--a change that would seem to represent social progress.

High School Curriculum. About the same percentage of students in the general, academic, and vocational-technical curriculums attended college in 1972 as in 1960 (see Table 20). However, the change in attendance in different high school courses has resulted in modest changes in the academic background of the college population. Just over 73 percent of the college students were enrolled in the academic course in high school.

Table 20. College Entrants Grouped by High School Curriculum

	College entrants			
High school curriculum	Estimated N*	% of stratum	% of cohort	Reading mean
General				
1960	95	24.8	14.3	10.4
1972	267	27.9	20.5	10.2
Academic				
1960	487	68.0	73.4	13.7
1972	937	68.0	72.0	12.6
Vocational-technical				
1960	83	13.5	12.5	10.6
1972	90	13.3	6.9	9.1
No response				
1960	57	38.1		12.3
1972				
Total				
1960	721	38.7	100.0	12.8
1972	1301	43.1	100.0	11.9

*In thousands

Those from general curriculums increased slightly at the expense of vocational-technical curriculums.

The reading test scores for college entrants who were enrolled in the general curriculum in high school showed very little decline from 1960 to 1972. College entrants who had taken vocational-technical curriculums showed a relatively large decline. On the whole, the results do not appear to contribute to an understanding of the score decline.

Expected College Major. Both in 1960 and 1972, students who expected to major in science or mathematics showed a higher proportion who actually entered college than any other group (see Table 21). In 1972, the percentage reached 86 percent. They were followed by students who expected to major in engineering (80 percent), then social sciences and humanities (77 percent), "other matching fields" (75 percent), and the students whose major fields could not be matched in the two studies (55 percent). The increases in the percentages from 1960 to 1972 followed a roughly reverse pattern. The subgroups with the highest 1960 percentages (science, mathematics, social sciences, and humanities) display the least increase, whereas engineering, other fields, and the "no matching field" group show greater gain. This may reflect in part a kind of ceiling effect: other things being equal, if nearly all

Table 21. College Entrants Grouped by College Major Choice

	College entrants			
Expected college major	Estimated N*	% of stratum	% of cohort	Reading mean
Science, mathematics				
1960	121	75.9	19.3	13.9
1972	181	85.6	15.6	13.5
Social sciences, humanities				
1960	130	74.3	20.6	13.6
1972	349	76.9	30.2	12.8
Engineering				
1960	87	63.2	13.8	13.9
1972	61	79.7	5.3	12.3
Other matching fields				
1960	246	53.6	39.0	12.2
1972	506	74.9	43.8	11.6
Nonmatching fields**				
1960	49	40.0	7.8	11.8
1972	42	55.1	3.6	10.9
No college plans; no response				
1960	92	11.3		12.0
1972	145	9.5		8.7
Total				
1960	721	38.7	100.0	12.8
1972	1301	43.1	100.0	11.9

*In thousands

**Fields that were listed in one survey but had no counterpart in the other survey

members of a subgroup go on to college, an increase in the percentage is less likely than it is for the groups with smaller percentages.

College entrants who expected to major in science or mathematics showed a relatively small score decline, but those who expected to enroll in engineering showed a relatively large decline. The latter decline was accompanied by a marked reduction in the percentage of students choosing engineering (14 percent of the cohort in 1960 and 5 percent in 1972). Although these results cannot readily be related to the score decline, it may be more than coincidental that engineering careers enjoyed great prominence during the post-Sputnik period when talent searching and talent loss were of great concern to higher education.

Summary of Changes for College Entrants. We conclude, therefore, that colleges are attracting more women students and students from nearly all types of families, particularly students from lower socioeconomic levels. This shift is more than offset, however, by changes in the educational and occupational status of parents, so that the colleges are even more heavily populated by high socioeconomic groups. As for the reading scores, the most striking finding, again, is the uniformity with which nearly all subgroups display a decline. A few exceptions were noted, but it appears likely that their combined effect would account for only a fraction of the observed decline.

Changes in the SAT Candidate Group

SAT scores have been widely used as a kind of social indicator, and some writers have interpreted the decline in SAT means as the consequence of a general decline in the scholastic ability of high school students. Such an interpretation need not be correct, because the decline in average SAT scores could simply represent a change in the population taking the SAT examinations.

As shown in Table 22, the most obvious change in the SAT population is the great increase in the proportion of students taking the test. In 1960, only 18 percent of the high school population took the SAT, whereas 31 percent were tested in 1972, an increase far greater than the increase in college attendance.

Two points that may affect the interpretation of the results deserve consideration. First, the actual number of students who took the SAT is approximately 3,500 for TALENT and approximately 4,500 for NLS. Although these samples are quite large, sampling error clearly plays a larger role, particularly in differences between means, for students who took the SAT than for high school seniors or college entrants.

The second complication in interpreting SAT-verbal means arises if the shift in the SAT score scale between 1960 and 1972 is taken into account. The results obtained in the Modu and Stern study (1976) and in the present study indicate that such a shift has occurred. Although the figures shown for SAT-verbal means are actual rather than adjusted, the Modu and Stern results will be taken into account when necessary in discussing the results. The reader who wishes to adjust 1972 means in order to obtain a more precise evaluation of particular 1972 means will find the following table useful.

If SAT score for 1972 group is	Subtract
563 - 587	5
538 - 562	6
513 - 537	7
476 - 512	8
438 - 475	9
413 - 437	10
388 - 412	11
363 - 387	12
350 - 362	13

Table 22. Students Who Took the SAT Grouped by Age

	SAT takers							
Age	N*	% of stratum	% of cohort	Reading mean	SAT-verbal		SAT-mathematical	
					Mean	S.D.	Mean	S.D.
17½ or less								
1960	74	27.2	22.8	14.7	487	107	509	109
1972	74	44.8	7.9	13.2	480	115	503	119
17½+ to 18								
1960	139	19.3	43.0	14.4	478	108	498	112
1972	404	34.5	43.3	12.6	457	104	492	110
18+ to 18½								
1960	91	16.5	27.9	14.0	467	107	490	111
1972	374	32.2	40.0	12.4	454	107	485	112
18½+ to 19								
1960	14	8.4	4.3	12.6	437	109	474	99
1972	63	20.2	6.8	11.3	424	110	455	113
Above 19								
1960	6	5.9	2.0	11.1	428	114	445	102
1972	18	10.0	2.0	8.4	353	104	393	109
No response								
1960	2	5.4		12.4	469	109	485	84
1972	1	5.7		7.3	342	42	337	57
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	936	31.0	100.0	12.4	453	108	485	113

*In thousands

Age. There has been a marked increase in the percentage of students taking the SAT at all age levels (see Table 22). In both the 1960 and 1972 data, increasing age is associated with a reduced probability of taking the SAT. All age groups declined in mean reading ability and in SAT means.

For SAT candidates, as for high school seniors and for college entrants, there has been a marked reduction in the percentage of students less than 17½ years old and a roughly corresponding increase in the percentage of students in the 18-to-18½-year-old group. In 1960, 23 percent of the SAT candidates were 17½ or younger; in 1972, only 8 percent were in this age group. This shift may have made a minor contribution to the score decline. If the reduction in the percentage of young SAT takers resulted from changed attitudes toward acceleration, however, the explanatory value of age shifts is weakened, as noted in the discussion of results for high school seniors.

For SAT-verbal scores, the youngest group shows an unusually small score decline--7 points as compared with 21 points for the overall group. The difference for reading scores is less pronounced--1.5 points as compared with 1.8 points.

Sex. Given a substantial increase for both sexes in the proportion taking the SAT, the increase for women is markedly greater. Women constitute 50 percent of the 1972 group as compared with 43 percent of the 1960 group (see Table 23). Although both males and females dropped in reading ability and in SAT means, the declines for females are somewhat larger. In 1960 the women had higher reading and SAT-verbal scores, presumably because of higher selectivity.

Interestingly, for high school seniors and college entrants the males showed somewhat greater score declines than females.

Parent's Education and Occupation. The SAT is taken by a larger proportion of students for every breakdown of education and occupation for both parents. The increase is especially noticeable at the lower socioeconomic levels where, for example, students whose fathers did not graduate from high school jumped from 10 percent to 21 percent (see Table 24), and the proportion for students whose father had a blue collar job jumped from 10 percent to 26 percent (see Table 26). Clearly, more students of all backgrounds are taking the SAT. As it happens, however, the *distribution* of parents of students who took the SAT has not changed very much. The proportion of SAT candidates whose fathers went to college stayed at about 50 percent of the fathers whose educational background was reported (see Table 24), while the proportion with fathers who did not graduate from high school declined from 25 percent in 1960 to 19 percent in 1972.

For parent's education, a similar pattern occurs for both the reading test and the verbal sections of the SAT. Students whose fathers had post-high-school education showed less than the average amount of decline, and students whose fathers had not attended college showed greater than average declines on both tests. A similar pattern holds for mother's education (see Table 25). For both fathers and mothers, the increase in the percentage who were high school graduates, and the corresponding decrease in the percentage without high school diplomas, might be expected to diminish rather than increase the extent of score decline.

Table 23. Students Who Took the SAT Grouped by Sex

	SAT takers							
					SAT-verbal		SAT-mathematical	
Sex	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
Males								
1960	187	20.5	57.3	14.1	471	109	520	111
1972	472	31.4	50.5	12.3	454	106	506	114
Females								
1960	140	14.6	42.7	14.4	478	107	465	102
1972	463	30.7	49.5	12.4	452	110	464	108
No response								
1960								
1972		10.8		15.0	600		520	
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	936	31.0	100.0	12.4	453	108	485	113

*In thousands

Table 24. Students Who Took the SAT Grouped by Father's Education

Father's education	SAT takers N*	% of stratum	% of cohort	Reading mean	SAT-verbal Mean	SAT-verbal S.D.	SAT-mathematical Mean	SAT-mathematical S.D.
Post high school								
1960	143	35.5	50.3	14.8	495	107	511	111
1972	435	48.5	49.3	13.3	484	106	517	109
High school graduate								
1960	70	18.5	24.5	14.1	462	105	492	109
1972	280	30.7	31.8	12.0	438	101	463	106
Less than high school graduate								
1960	72	9.6	25.2	13.5	449	107	481	110
1972	167	21.0	18.9	11.5	423	99	466	109
Don't know; blank								
1960	42	12.6		13.7	466	110	478	106
1972	53	12.9		9.7	379	100	402	108
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	935	31.0	100.0	12.4	453	108	485	113

*In thousands

Table 25. Students Who Took the SAT Grouped by Mother's Education

	SAT takers							
					SAT-verbal		SAT-mathematical	
Mother's education	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
Post high school								
1960	127	34.6	43.5	14.8	496	107	513	112
1972	367	49.3	41.3	13.2	485	108	517	111
High school graduate								
1960	100	18.4	34.3	14.1	465	105	494	107
1972	383	32.1	43.0	12.3	443	101	477	108
Less than high school graduate								
1960	65	9.5	22.2	13.5	443	103	474	108
1972	140	19.6	15.7	11.4	421	101	450	106
Don't know; blank								
1960	35	13.1		13.9	476	112	484	109
1972	45	12.5		9.4	385	104	404	112
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	935	31.0	100.0	12.4	453	108	485	113

*In thousands

For father's occupation (see Table 26), there seems to be little difference among the three major groups on the verbal sections of the SAT, but students whose fathers held white collar positions show a slightly greater decline. Students in the "Don't know; blank" category show a decidedly greater than average decline on both tests. For mother's occupation (see Table 27), students whose mothers were engaged in professional or managerial occupations or who were homemakers showed a slightly greater score decline than students whose mothers were engaged in white collar or blue collar occupations, again on both tests. Although the differences on the verbal sections of the SAT are reduced when the NLS means are adjusted for scale drift, the pattern remains the same.

The interpretation of the result for homemakers is complicated by the fact that, in the TALENT questionnaire, students were told not to check "Housewife" if their mother had worked for pay in the last three years, while NLS imposed no such restriction.

Family Configuration. In general, most of our observations in regard to the families of all college entrants are applicable to students who took the SAT. As shown in Table 28, in 1960 a higher proportion of the higher ability configurations took the SAT, and similarly, but to a lesser extent, in 1972. In other words, family configuration became, in 1972, a less important determinant of taking the SAT, although we still find a somewhat smaller proportion of students from large families doing so.

For the TALENT subjects, the mean reading scores of the various strata of students who took the SAT follow the pattern predicted by the confluence model in three ways: (1) within families the first-born children have scores equal to or higher than the scores of later born children; (2) only children have a mean that is less than the mean for the first born in families of four; and (3) the lowest mean is the mean for all children in families of six or more. The separation of the various configurations is, however, substantially less than it is for the high school senior population. This suggests that self-selection strongly attenuated the effect of family configuration. In any case, the effects in question are small, for again we find that within each stratum the mean reading scores decreased by an amount that in most cases approximates the overall decline.

The SAT-verbal means exhibit patterns similar to the reading score means. For both TALENT and NLS, those who are only children have lower SAT means (both verbal and mathematical) than the first born of families of two, three, and four and, by and large, the subjects in larger families have lower scores. There are many exceptions to the expected patterns, however, possibly because of sampling error. (In one case the mean is based on only 56 actual cases.) Furthermore, we find the same attenuation of the effect that seems to exist for the reading score; the means for the various configurations simply do not display the same spread as those found for the reading scores of the high school population.

In order to estimate the magnitude of the effect of the change in family configuration from 1960 to 1972, we asked what the 1972 mean SAT-verbal score would have been if nothing changed from 1960 to 1972 except for the composition of the SAT cohort. The weighting of each

Table 26. Students Who Took the SAT Grouped by Father's Occupation

	SAT takers							
					SAT-verbal		SAT-mathematical	
Father's occupation	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
Professional, managerial								
1960	118	37.8	41.8	14.8	497	107	514	111
1972	361	49.3	42.9	13.3	482	107	516	109
White collar								
1960	70	27.1	24.8	14.4	470	104	493	109
1972	170	39.0	20.2	12.4	455	106	487	110
Blue collar								
1960	94	9.6	33.4	13.6	449	106	485	108
1972	310	25.6	36.9	12.0	437	99	472	107
Don't know; blank								
1960	45	14.2		13.8	472	111	480	109
1972	94	14.8		10.0	391	106	409	106
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	935	31.0	100.0	12.4	453	108	485	113

*In thousands

Table 27. Students Who Took the SAT Grouped by Mother's Occupation

Mother's occupation	SAT takers							
	N*	% of stratum	% of cohort	Reading mean	SAT-verbal		SAT-mathematical	
					Mean	S.D.	Mean	S.D.
Professional, managerial								
1960	38	27.6	13.1	14.9	503	111	516	107
1972	132	46.6	15.7	13.1	485	111	510	111
White collar								
1960	60	25.2	20.7	14.4	476	106	498	114
1972	201	38.2	23.9	12.8	464	99	498	107
Blue collar								
1960	32	9.7	11.0	13.5	444	106	471	106
1972	79	26.2	9.4	12.2	436	100	478	109
Homemaker								
1960	161	17.9	55.2	14.2	473	107	499	111
1972	429	31.4	51.0	12.5	455	107	488	113
Don't know; blank								
1960	35	13.5		13.9	471	109	482	105
1972	94	17.6		10.0	394	107	417	106
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	936	31.0	100.0	12.4	453	108	485	113

*In thousands

Table 28. Students Who Took the SAT Grouped by Family Configuration

	SAT takers							
					SAT-verbal		SAT-mathematical	
	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
Only child								
1960	34	24.6	11.7	14.2	471	109	489	113
1972	43	44.7	5.0	12.9	463	111	487	111
First of two								
1960	66	27.9	22.6	14.5	476	106	497	113
1972	93	42.0	10.8	12.8	467	105	491	111
Second of two								
1960	36	23.1	12.3	14.3	478	108	500	112
1972	109	41.5	12.7	12.5	457	102	485	115
First of three								
1960	36	21.5	12.3	14.7	478	110	501	110
1972	84	39.2	9.8	12.9	468	111	497	109
Second of three								
1960	21	18.8	7.1	14.2	466	101	503	107
1972	89	37.8	10.3	12.5	457	111	494	113
Third of three								
1960	13	17.4	4.6	14.3	479	111	504	109
1972	71	34.7	8.2	12.1	448	106	478	111
First of four								
1960	19	20.3	6.6	14.4	491	106	511	111
1972	43	32.8	5.0	12.7	466	112	497	114

(Continued)

Table 28 (Continued)

	SAT takers							
					SAT-verbal		SAT-mathematical	
	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
Second of four								
1960	11	15.7	3.8	13.9	461	105	503	120
1972	43	30.1	5.1	12.4	451	107	495	110
Third of four								
1960	7	14.8	2.4	14.4	474	98	494	92
1972	49	32.6	5.7	12.7	460	100	491	112
Fourth of four								
1960	5	13.3	1.8	14.2	470	107	488	89
1972	25	26.1	2.9	12.7	476	112	506	125
First of five								
1960	7	15.3	2.5	14.3	483	114	517	120
1972	19	27.2	2.3	11.9	444	103	508	107
Second of five								
1960	5	13.0	1.6	14.3	459	112	493	99
1972	22	29.9	2.5	12.3	446	108	490	108
Third of five								
1960	3	11.2	1.1	13.9	476	122	496	92
1972	22	25.9	2.5	12.3	443	102	492	102
Fourth of five								
1960	2	7.3	0.6	13.7	450	92	485	95
1972	14	24.9	1.6	12.3	432	98	482	108

Table 28 (Continued)

	SAT takers							
					SAT-verbal		SAT-mathematical	
	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
Fifth of five								
1960	2	10.0	0.8	13.8	456	92	483	89
1972	11	28.8	1.3	13.3	464	96	482	117
Six or more children								
1960	24	7.8	8.3	12.8	453	116	478	112
1972	121	21.3	14.1	12.1	437	108	476	112
No data								
1960	35	13.5	0.0	14.2	480	108	487	105
1972	79	21.1	0.0	10.9	428	109	447	114
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	935	31.0	100.0	12.4	453	108	485	113

*In thousands

Table 29. Students Who Took the SAT Grouped by High School Curriculum

	SAT takers							
					SAT-verbal		SAT-mathematical	
High school curriculum	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
General								
1960	19	4.9	6.2	11.9	423	109	442	106
1972	110	11.5	11.8	10.7	397	97	434	104
Academic								
1960	267	37.3	88.3	14.5	480	106	504	109
1972	783	56.8	83.7	12.8	466	105	498	110
Vocational-technical								
1960	17	2.7	5.5	11.5	410	105	428	98
1972	43	6.3	4.6	9.3	359	82	379	89
No response								
1960	24	16.0		14.5	495	109	497	108
1972								
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	936	31.0	100.0	12.4	453	108	485	113

*In thousands

1960 mean by the 1972 "% of cohort" produced a new overall mean of 472. This is the best estimate of the change in SAT-verbal scores that would have occurred if the mean scores within groups had remained the same but the distribution of SAT candidates with respect to family configuration types had changed (as they did). The estimated mean is 2 scale points less than the 1960 overall mean of 474. Thus, this particular statistical technique suggests that changes in family configuration contributed only a small amount to the decline in SAT-verbal scores from 1960 to 1972.

High School Curriculum. Although the percentages of students in different curriculums who went on to college remained unusually stable from 1960 to 1972, the percentages who took the SAT consistently increased (see Table 29). The percentage of general students taking the SAT increased markedly, from 5 percent to 12 percent, as did the percentage of vocational-technical students, from 3 percent to 6 percent. When the changes in the total numbers enrolled in the various curriculums are considered, we find that the composition of students who took the SAT changed from 6 percent to 12 percent for general students, from 88 percent to 84 percent for college preparatory students, with almost no change for vocational-technical students.

When students are grouped by curriculum, the academic group showed a noticeably smaller score decline than students in vocational-technical curriculums among students who took the SAT, both on that test and on the reading test.

Expected College Major. As with high school curriculum, the increase in the percentage of each college major group taking the SAT increased much more than the percentage entering college (see Table 30). The most dramatic increase was in the case of the students who expected to major in "Other matching fields" (for example, business, education, and nursing). The number of such students increased from 98,000 to 319,000, or from 22 percent to 47 percent of the cohort. The other subgroups increased also but not to the same degree. When changes in the number in each subgroup are considered, we find the following substantial changes in the composition of students taking the SAT: (1) an increase of 10 percent in the students planning to major in social sciences and the humanities, (2) a decrease of 11 percent in engineering majors, and (3) an increase of 8 percent in other majors.

Prospective science and mathematics majors showed a slight gain in SAT-verbal scores and a relatively small loss in reading scores. Students whose major field was classified in "Other matching fields" showed relatively small losses in SAT-verbal and reading scores.

College Attendance. A question of considerable interest concerns changing patterns of college attendance among test takers. Table 31 indicates a substantial change. In 1960, 77 percent of students who took the SAT attended four-year colleges; in 1972, the percentage was only 60 percent. Part of this difference is attributable to the increase in two-year college attendance, which increased from 6 percent of the students who took the SAT to 15 percent. Rather surprisingly, the percentage who were not attending college increased from 8 percent to 16 percent. Because the four-year college group was substantially higher in test performance than the groups that increased in their proportion of the population (about 67 scaled score points higher in 1960 and about 70 points higher in 1972 on the verbal sections of the

Table 30. Students Who Took the SAT Grouped by College Major Choice

	SAT takers							
					SAT-verbal		SAT-mathematical	
Expected college major	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
Science, mathematics								
1960	60	37.3	20.8	14.9	499	109	544	113
1972	137	64.9	17.9	13.9	503	106	560	106
Social sciences, humanities								
1960	64	36.9	22.5	15.1	509	106	492	101
1972	245	54.0	32.0	13.3	486	105	496	106
Engineering								
1960	46	33.4	16.0	14.8	485	100	566	99
1972	40	52.2	5.3	13.1	470	118	558	108
Other matching fields								
1960	98	21.5	34.3	13.3	437	101	450	96
1972	319	47.2	41.8	12.0	436	96	471	100
Nonmatching fields								
1960	18	14.8	6.3	13.5	438	98	464	94
1972	23	29.4	3.0	11.2	424	81	487	104
No college plans; blank								
1960	40	4.9		13.7	474	110	479	105
1972	172	11.3		10.6	399	102	420	106
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	936	31.0	100.0	12.4	453	108	485	113

*In thousands

Table 31. Students Who Took the SAT Grouped by College Attendance

	SAT takers							
					SAT-verbal		SAT-mathematical	
College attendance	N*	% of stratum	% of cohort	Reading mean	Mean	S.D.	Mean	S.D.
Four-year college, full-time student								
1960	197	44.2	77.1	14.9	495	104	516	108
1972	539	64.5	60.2	13.2	483	106	520	108
Two-year college, full-time student								
1960	15	18.0	6.0	12.9	426	92	443	100
1972	134	35.5	15.0	11.1	411	88	438	95
Vocational or other school								
1960	22	10.4	8.6	13.5	430	104	464	106
1972	74	22.5	8.3	11.2	412	96	435	103
Not in school**								
1960	21	4.8	8.4	13.0	434	99	457	103
1972	148	11.6	16.5	11.3	412	101	434	102
Unclassified								
1960	71	10.4		13.2	451	113	474	108
1972	40	19.9		11.1	425	109	460	121
Total								
1960	326	17.5	100.0	14.2	474	108	496	110
1972	936	31.0	100.0	12.4	453	108	485	113

*In thousands
**Includes part-time college students

SAT), these results describe a population change that does contribute appreciably to the decline in SAT scores.

SUMMARY

In general, we observe a great increase in the proportion of high school seniors who take the SAT, as a result of the more rapid growth in the SAT candidate volume from 1960 to 1972 than the growth in the number of high school graduates. Of greater interest are the results involving changes in the *composition* of the SAT population brought about by the increases and decreases reported above. When shifts in the number of students in each subgroup are considered, we find the following changes in the SAT population:

1. A marked increase in the proportion of women candidates
2. A small increase (5 percent) in the proportion of "general" students and an equal decrease in the proportion of college preparatory students
3. An increase in the proportion of students from larger families
4. Increases in the proportion of students planning to major in the social sciences and humanities and certain undergraduate career programs (for example, business, education, and nursing) and a decrease in engineering majors
5. A substantial decrease in the *proportion* of SAT takers who attended four-year colleges

On both tests, almost every subgroup of students who took the SAT showed some decline in scores. Among the groups that showed a slight increase or a relatively small decline were students younger than 17½ years, students whose parents had education beyond high school, students who expected to major in science or mathematics in college, or in some field other than liberal arts and engineering, students whose mothers were employed in white collar or blue collar occupations, and students who entered four-year colleges. Almost without exception, the decline in scores for these groups was more than half as large as the average decline. One change appears to have made an appreciable contribution to the SAT score decline: a substantial decrease in the proportion of candidates who enter four-year colleges.

REFERENCES

Angoff, W. H., "Scales, Norms and Equivalent Scores," *in* R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, D.C.: American Council on Education, 1971

Beaton, A. E., and Tukey, J. W., "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data." *Technometrics*, 1974, No. 16, pp. 147-185

Bianchini, J. C., and Loret, P. G., *Anchor Test Study: Final Report*. Princeton, N.J.: Educational Testing Service, 1972

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., and York, R. L., *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office, 1966

College Entrance Examination Board, *Report of the Commission on Tests. I. Righting the Balance*. N.Y.: College Entrance Examination Board, 1970

Flanagan, J. C., "Changes in School Levels of Achievement: Project TALENT Ten and Fifteen Year Retests." *Educational Researcher*, 1976, Vol. 5, No. 8, pp. 9-12

Flanagan, J. C., and Jung, S. M., *Progress in Education: A Sample Survey (1960-1970)*. Palo Alto, Calif.: American Institutes for Research, 1971

Flanagan, J. C., Dailey, J. T., Shaycoft, M. F., Gorham, W. A., Orr, D. B., and Goldberg, I., *Design for a Study of American Youth*. Boston: Houghton Mifflin, 1962

Flanagan, J. C., Davis, F. B., Dailey, J. T., Shaycoft, M. F., Orr, D. B., Goldberg, I., and Neyman, F. A., Jr., *The American High School Student* (Final report to the U.S. Office of Education Cooperative Research Project No. 635). Pittsburgh, Pa.: Project TALENT Office, University of Pittsburgh, 1964

Frankel, M. M., and Beamer, J. F., *Projections of Educational Statistics to 1982-83* (1973 ed.). Washington, D.C.: U.S. Government Printing Office, 1974

Harnischfeger, A., and Wiley, D. E., *Achievement Test Score Decline: Do We Need to Worry?* Chicago, Ill.: ML-Group for Policy Studies in Education, CEMREL, Inc., 1975

Hilton, T. L., Rhett, H., Broudy, I. L., Bower, C., Carter, M. M., Creech, F. R., and Echternacht, G. J., *The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972* (Final report to the U.S. Department of Health, Education, and Welfare, National Center for Education Statistics, Contract No. OEC-0-72-0903). Princeton, N.J.: Educational Testing Service, 1973

Jackson, R., and Schrader, W. B., *Verbal and Mathematical Ability of High School Juniors in 1974: A Norms Study of PSAT/NMSQT*. College Entrance Examination Board Research and Development Report 76-77, No. 2. Princeton, N.J.: Educational Testing Service, 1976

Kish, L., and Frankel, M. R., "Balanced Repeated Replication for Standard Errors." *Journal of the American Statistical Association*, 1970, Vol. 65, No. 65, pp. 1071-1094

Kish, L., and Frankel, M. R., "Inference from Complex Samples." *Journal of the Royal Statistical Society, Series B*, 1974, No. 36, pp. 1-22

McCarthy, P. J., "Pseudo-Replication: Half-Samples." *Review of the International Statistical Institute*, 1969, No. 37, pp. 305-325

Modu, C. C., and Stern, J., *The Stability of the SAT-Verbal Score Scale*. Unpublished report. Princeton, N.J.: Educational Testing Service, 1976

National Center for Education Statistics, U.S. Office of Education, *Directory of Public Elementary and Secondary Day Schools, 1968-69* (OE-20126, Vols. I-IV). Washington, D.C.: U.S. Government Printing Office, 1970

National Center for Education Statistics, U.S. Office of Education, *Directory of Nonpublic Elementary and Secondary Day Schools, 1968-69* (OE-20127, Vol. V). Washington, D.C.: U.S. Government Printing Office, 1970

Perry, D., and Swanson, E. O., "Decline in Minnesota College Aptitude Test Scores," *in* J. Fasold, M. Hall, and J. Impara (Eds.), *Decline* *in Standardized Test Scores: A Widespread Phenomenon*. Salem, Ore.: Oregon Department of Education, 1974

Plackett, R. L., and Burman, J. P., "The Design of Optimum Multifactorial Experiments." *Biometrika*, 1943-1945, Vol. 33, pp. 305-325

Schrader, W. B., and Hilton, T. L., *Educational Attainment of American High School Seniors in 1960, 1965, and 1972: Feasibility Study* (Project NIE-G-74-0050) (ETS PR-75-13). Princeton, N. J.: Educational Testing Service, 1975

Simon, K. A., and Frankel, M. M., *Projections of Educational Statistics to 1981-82* (1972 ed.). Washington, D.C.: U.S. Government Printing Office, 1973

Stricker, L. J., *Dimensions of Social Stratification for Whites and Blacks: The Toledo Study* (ETS RB-76-30). Princeton, N.J.: Educational Testing Service, 1976

Thorndike, R. L., "Concepts of Cultural Fairness." *Journal of Educational Measurement*, 1971, No. 8, pp. 63-70

U.S. Bureau of the Census, *County and City Data Book*, 1972 (A statistical abstract supplement). Washington, D.C.: U.S. Government Printing Office, 1972

Zajonc, R. B., "Family Configuration and Intelligence." *Science*, 1976, Vol. 19, pp. 227-236

Zajonc, R. B., and Markus, G. B., "Birth Order and Intellectual Development." *Psychological Review*, 1975, Vol. 82, No. 1, pp. 74-88

APPENDIX A. EQUATING THE PROJECT TALENT AND NATIONAL LONGITUDINAL STUDY READING TESTS

Introduction

A careful survey of the extensive battery of tests used in Project TALENT in 1960 and the tests used in the NLS survey in 1972 indicated that the tests of reading comprehension and the mathematical tests were sufficiently similar to warrant formal study of their equivalence (Schrader and Hilton, 1975, pp. 68-70). When the available evidence on the tests was reviewed in the light of the present study design, however, the probability that the two mathematical tests could be properly equated seemed to us not high enough to warrant their inclusion. Accordingly, we focused our attention on the reading tests.

Equating Design

The equating design was based on Angoff's Design II (Angoff, 1971, pp. 573-576). In this design, the two tests to be equated are administered so that one random half of the equating sample takes the tests in one order and the other random half takes them in the reverse order.

Each participating school was asked to test only one class (preferably an English class composed mainly of seniors) and to administer the two tests at different class sessions. The order in which the tests were to be administered was specified in the instructions sent to each school.

Sample Design

For public schools in the 50 states and the District of Columbia, stratification was based on (1) geographical region and (2) whether or not the school was located in a Standard Metropolitan Statistical Area (SMSA). All private schools were included in a single stratum.

Enrollment data for public high school students (grades 9-12) for SMSAs and non-SMSAs by state were obtained from reports of the 1970 census (U.S. Bureau of the Census, 1972). In all, 11 strata approximately equal in high school enrollment were developed, as follows:

Stratum	Approximate enrollment (in thousands)
Metropolitan public schools:	
New England, New York	1,352
Other North Atlantic	1,125
Indiana, Ohio, Michigan	1,201
Other Great Lakes and Plains	1,178
Southeast	1,339
California	1,229
Other West and Southwest	1,250

(Continued)

Stratum	Approximate enrollment (in thousands)
Nonmetropolitan-public schools:	
North Atlantic, East North Central	1,438
Southeast.	1,492
West North Central, West, Southwest.	1,461
Private schools	
All regions.	1,411

Because the schedule for the study did not permit replacement of schools unable or unwilling to participate, we invited 176 schools (16 from each stratum) to administer the tests. Assuming that about half would participate, we expected to obtain data for about 88 schools for our actual equating.

The College Board/ETS Secondary School Master File, which includes more than 25,000 schools, constituted our basic list for school selection. The master file, reviewed monthly, includes public, independent, and parochial schools and is as comprehensive as possible. From this list we selected a random sample of 1,000 schools and assigned a random sequence number to each of them. Data needed for assigning schools to strata and to confirm that the school had a twelfth grade were obtained from various sources, particularly the SMSA list in the 1972 County and City Data Book (U.S. Bureau of the Census, 1972) and the five volumes of the directory of secondary schools, 1968-69 (National Center for Education Statistics, 1970). Within each stratum, we selected the 16 schools having the lowest sequence numbers. Two schools invited to participate in the 1974 PSAT/NMSQT Norms Study were replaced by schools in the same stratum having the next following sequence numbers. Altogether, we invited a total of 176 schools to participate. (Because of an error in assigning schools to strata, one of the 176 schools should have been replaced by a different school. However, the error was discovered too late to permit inviting a replacement school. The school that should not have been invited did not participate.)

Test Administration

Because the study was initiated relatively late in the academic year, we mailed invitations to participate on March 29 and asked the schools to test in late April or early May. All but one of the schools that did not reply were reached by telephone, and a total of 91 schools administered the tests.

We reproduced the tests from copies of the tests printed for the original survey. Insofar as possible, we left the directions to the students unchanged, but we prepared special directions for each order of testing, to minimize the risk that they would be given in the wrong order. We designated the TALENT test Form 1 and the NLS test Form 2 in schools that were to give TALENT first. For schools giving NLS first, we designated the NLS test Form A and the TALENT test Form B. We allowed schools to give makeup tests to absentees but did not use the scores for these students in the equating.

Each school was given a set of guidelines designed to provide an objective basis for selecting the class to be tested--guidelines based on the first letter of the last name of the teacher and on the time the class met. We did not collect data on how precisely the schools followed the guidelines.

Following the testing, we reviewed supervisors' reports and answer sheets. One school reported a serious mistiming and was excluded from the equating sample. Two schools had reversed the order of administration; their data were analyzed on the basis of the order in which the tests were actually given. We excluded a number of students whose testing date indicated that they had been given makeup tests.

Equating Sample

Our equating sample included only students who had taken both tests and who reported that they were in the twelfth grade. We excluded from the study two schools that had no usable data for twelfth-grade students. Table A1 shows the distribution of schools and students by strata and by testing orders. In the final equating sample, there were 47 schools that administered TALENT first and 41 that administered NLS first. In part, this difference in the number of schools for the two testing

Table A1. Distribution of Participating Schools and Students by Stratum and Testing Order

Stratum	TALENT-NLS No. of schools	Sequence No. of students[a]	NLS-TALENT No. of schools	Sequence No. of students[a]
Metropolitan public schools:				
New England, New York	6	94 (9)	4	75
Other North Atlantic.	4	86 (2)	4	84 (3)
Indiana, Ohio, Michigan . . .	5	73 (1)	5	81 (6)
Other Great Lakes and Plains.	3	49 (2)	3	51
Southeast	3	43 (3)	4	51 (1)
California.	5	97 (1)	3	59 (3)
Other West and Southwest. . .	2	33	4	89 (1)
Nonmetropolitan public schools:				
North Atlantic, East North Central	7	139	5	98 (1)
Southeast	4	75 (1)	3	65 (2)
West North Central, West, Southwest.	4	141 (6)	2	34 (4)
Private schools:				
All regions	4	94 (3)	4	96 (1)
Total.	47	924 (28)	41	783 (22)

a. Numbers in parentheses indicate the number of students excluded on the basis of robust regression analysis from the final equating sample.

orders arose because testing orders were assigned alternately within strata using a random start, after each school agreed to participate. As it turned out, each of the four strata that had an odd number of participants was assigned the TALENT-NLS sequence.

Each school was assigned a predetermined weight, so that the weighted sample would have the following characteristics:

1. The sum of the weighted frequencies for each stratum would be proportional to the number of students in that stratum in grades 9-12 in the 1970 census data.
2. Within each stratum, the sum of the weighted frequencies would be equal for each school.

Weights were determined separately for each testing order.

Characteristics of the Equating Sample

As part of the data collection for the equating study, we asked each student to complete a 10-item questionnaire, primarily to provide us with a means of comparing the characteristics of our equating sample with those of Project TALENT and NLS. From 98 percent to 100 percent of the twelfth-grade students in the sample completed the questionnaire, depending upon the item. Table A2 shows the percentage of the sample that responded to each item and also, when comparable data existed, the distribution of percentages for three other samples:

1. The 1960 TALENT sample of twelfth graders, as reported in the *American High School Student* (Flanagan et al., 1964)
2. The 1970 TALENT Sample Resurvey (Flanagan and Jung, 1971)
3. The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972, as reported in Hilton et al., 1973

These results are discussed in the following paragraphs.

First, the 100 percent figure for the twelfth-grade students in the equating sample merely reflects the fact that for equating and for these comparisons all students below grade 12 were excluded from the sample.

The appreciably smaller proportion of males in the equating sample probably resulted from the instruction to the schools to select the participating classrooms from those taught by "teachers of twelfth-grade English (or closely related subjects like communications or journalism)." We assume that a somewhat smaller proportion of men than women are enrolled in senior English courses. Also, although schools were permitted to conduct makeup test administrations for students who were absent on either or both testing days, results from these optional administrations were not included in the analyses. More boys than girls may have been absent.

The third item (question 23) indicates that the equating sample contained a disproportionately large number of academic or college preparatory students. This, again, probably resulted from selecting the sample classrooms from English classes. Although most seniors take one subject taught by the English department, the probability of a student's taking more than one such subject is no doubt higher for students in academic or college preparatory programs. Even if one assumes a tendency for students to overstate their standing, question 24 suggests that the sample students tended to be above average. Fourteen percent reported they were in the top 5 percent of their class, and only 3 percent acknowledged they were in the low quarter.

Table A2. Percentages for Equating Study Questionnaire Responses and Comparisons with Project TALENT and NLS Survey Results

	% TALENT 1960[a]	% TALENT 1970[b]	% NLS 1972[c]	% Equating sample[d]
21. What grade are you now in?				
(A) 12	100.0		100.0	100.0
(B) 11		100.0		
(C) 10 or lower.				
	100.0	100.0	100.0	100.0
22. What is your sex?				
(A) Male	49.3	47.7	49.7	43.6
(B) Female	50.7	52.3	50.4	55.9
Omitted.				.5
	100.0	100.0	100.1	100.0
23. Which of the following best describes your present high school program?				
(A) General.	20.6	25.9	30.6	32.2
(B) Academic or college preparatory.	38.4	44.3	45.9	54.3
(C) Commercial or business (including distributive education), vocational or technical (including health occupations and home economics occupations), agriculture or a program very different from above	32.9	29.9	23.5	13.0
Omitted or no data	8.1			.6
	100.0	100.1	100.0	100.1
24. How would you compare your academic achievement, as measured by grades or rank, with that of the other students in your high school class?				
(A) Top 5%		No comparable items		14.1
(B) Top quarter but not top 5% .				34.5
(C) Second quarter (from the top)				35.9
(D) Third quarter (from the top)				10.8
(E) Low quarter.				3.1
Omitted.				1.6
				100.0

(Continued)

Table A2. (Continued)

	% TALENT 1960[a]	% TALENT 1970[b]	% NLS 1972[c]	% Equating sample[d]
25. Which of the following best describes your plans for next year?				
(A) Get a full-time job or join the military service.			29.2	18.1
(B) Attend a 4-year college. .	42.0[e]		33.6	49.3
(C) Attend a 2-year college. .	10.0		16.3	19.9
(D) Become a full-time homemaker.			2.8	.9
(E) Other.			18.2	11.0
Omitted.				.7
			100.1	99.9
26. Did you take the College Board Scholastic Aptitude Test?				
(A) Yes.			29.8	45.2
(B) No			70.2	54.1
Omitted.				.7
			100.0	100.0
27. Did you take the American College Test (ACT)?				
(A) Yes.			19.8	42.3
(B) No			80.2	56.8
Omitted.				.9
			100.0	100.0
28. How many books have you read (not including those required for school) in the past 12 months? Don't count magazines or comic books.				
(A) None	11.7	9.6		8.5
(B) 1-5.	40.4	39.6		42.6
(C) 6-10	19.6	22.4		23.5
(D) 11-15.	10.2	11.5		10.4
(E) 16 or more	18.2	16.9		14.4
Omitted.				.6
	100.1	100.0		100.0
29. Did your father attend college?				
(A) Yes.	19.1[f]		30.5	40.0
(B) No			69.5	58.7
Omitted.				1.3
			100.0	100.0

(Continued)

Table A2. (Continued)

	% TALENT 1960[a]	% TALENT 1970[b]	% NLS 1972[c]	% Equating sample[d]
30. Did your mother attend college?				
(A) Yes.	15.0[g]		21.9	32.1
(B) No			78.1	66.9
Omitted.				1.0
			100.0	100.0

a. Both TALENT samples are weighted.
b. From Flanagan and Jung (1971).
c. Weighted by means of base-year weights for students completing the student questionnaire.
d. Weighted in accordance with sample design.
e. From item 302 of the TALENT student questionnaire. Other responses were not relevant.
f. From item 218 of the TALENT student questionnaire. Other responses were not relevant.
g. From item 219 of the TALENT student questionnaire. Other responses were not relevant.

As one might expect from the relatively high proportion of sample subjects in the college preparatory program, the proportion who reported that they planned to attend college exceeds by about 20 percent the proportion of NLS subjects who planned to attend college, and again, as would be expected, a higher proportion of the sample subjects took either the SAT or ACT.

Regarding the number of books that the sample subjects had read in the past 12 months (question 28), the distribution of frequencies is remarkably similar to that in the 1960 and 1970 TALENT samples. Why these results should be so invariant is not clear.

Lastly, approximately 10 percent more of the students reported that their fathers attended college than did the students in the NLS sample and, similarly, for college attendance by mothers.

Another relevant descriptor is the performance of the equating sample on the two reading tests in question. The equating sample was clearly superior. The students who took NLS first had a mean of 11.71 and a standard deviation of 4.80, compared with a mean of 9.65 and a standard deviation of 5.05 for the 1972 NLS sample (Hilton et al., 1973). Similarly, the students in the equating sample who took TALENT first had a mean of 34.5 and a standard deviation of 9.9, compared with a mean of 33.3 and a standard deviation of 10.1 for the twelfth-grade sample of TALENT students (Flanagan et al., 1964). Thus, it appears that the equating sample subjects are an above average group of students in ability, educational aspirations, and socioeconomic background.

Equating Method

Essentially, the equating method used called for equipercentile equating of TALENT and NLS scores for each testing order separately and for averaging the results. It was decided to express TALENT scores on the NLS scale in this study, so that for each TALENT raw score the average of the two NLS equivalent scores defined the line or relation. This method of equating was found to be effective in the Anchor Test Study (Bianchini and Loret, 1972).

The program used for the equipercentile equating is based on linear interpolation, using the midpercentile for each raw score on each test as the basis for interpolating. Results are calculated to one decimal place. A special problem in the present study arose because NLS scores were calculated by the conventional formula for correcting for chance success. After some exploration of ways to deal with this problem, it was decided to use the alternative scoring formula: rights plus one-fifth omits, and round the result to the nearest integer. After the equating had been completed, each NLS score was transformed to the corresponding score for conventional formula scoring. For a 20-item test composed of 5-choice items, scores on the alternative scale can be expressed on the conventional scale by multiplying the score by 1.25 and subtracting 5. For example, a student who had 11 questions right, 4 wrong, and 5 omitted would receive a score of 12 on the "rights plus 1/5 omits" scale and a score of 10 on the "rights minus 1/4 wrong" scale.

One other point about the equating process deserves mention. As it turned out, data were insufficient to determine the equating line for the lowest TALENT raw scores. It was decided to define the scale in this range by a linear extrapolation based on the NLS equivalent scores for the seven lowest scores for which equivalent scores could be calculated.

Equating of Attenuated Scores

Any attempt to equate scores on two tests must take into account the extent to which the tests differ in what they measure, in their reliability, and in their difficulty level. When the tests whose scores are to be equated are designed by the same authors to measure the same abilities and when they are matched with respect to format and difficulty level, appropriate equating methods are likely to produce interchangeable scores. On the other hand, when the tests differ in format and difficulty level and are prepared by different authors for somewhat different purposes, the problem of parallelism is formidable. Because the tests differed in difficulty level, it was recognized that the line of relation should be curvilinear. As Angoff (1971, pp. 567-568) makes clear, there are serious logical and statistical problems in equating tests that differ in reliability. There were definite reasons for believing that the TALENT and NLS reading tests differed in reliability, because the TALENT test included 48 items with a 30-minute time limit and the NLS test included 20 items with a 15-minute time limit. Data from the equating sample confirmed the expectation that the TALENT scores were substantially more reliable. An exploratory study was made of the possible use of Lord's Item Characteristic Curve approach to equating, but because item data were not available for the TALENT sample, it was concluded that an alternative approach would be needed. It was finally decided to attenuate the TALENT scores by adding random normal deviates to the observed scores. Because the equating was done solely to provide as accurate a comparison as possible between the TALENT and NLS scores for the present study, the modification of TALENT scores by introducing random error appeared to be an acceptable method for balancing out the difference in test reliability.

The correlation coefficients of TALENT scores and of NLS scores with SAT-verbal scores and the standard deviations of TALENT scores for 643 students in the 1976 equating sample who had taken the SAT were used in determining the amount of additional error variance to be included in the TALENT scores. The following equation provided the basis for determining the standard deviation of the random deviates to be added to TALENT scores:

$$r_{01} = \frac{\Sigma x_0 x_2}{\sigma_0 \sqrt{\sigma_2^2 + \sigma_r^2}}$$

In this equation,

0 = SAT-verbal
1 = NLS reading
2 = TALENT reading
σ_r^2 = variance of the random digits

The foregoing equation assumes that the covariance of the random digits with SAT-verbal and TALENT scores equals zero. The numerator may be written

$r_{02}\sigma_0\sigma_2$.

When the resulting equation is solved for σ_r^2, the result is:

$$\sigma_r^2 = \left(\frac{r_{02}^2}{r_{01}^2} - 1 \right) \sigma_2^2$$

When observed values for the sample of 643 students are substituted in this equation, the following equation is obtained:

$$\sigma_r^2 = \left(\frac{.7869^2}{.7357^2} - 1 \right) (7.2571)^2$$

and

$\sigma_r = 2.75$

Accordingly, the TALENT scores were modified by adding a random normal deviate to each. Normal deviates for this purpose were obtained by using random normal deviates with a mean of zero and a standard deviation of 1, multiplying each deviate by 2.75, and curtailing the distribution by limiting the range to values between -7 and +7. The modified TALENT scores were expressed as integers by discarding figures to the right of the decimal point.

As part of the exploratory study of the Item Characteristic Curve approach to equating, the combined equating sample was analyzed and 50 students whose performance on the two tests was markedly inconsistent were removed by means of the so-called robust regression procedure (Beaton and Tukey, 1974). The final equating used in the main analysis was based on the reduced sample. Table A3 shows the NLS score corresponding to each modified TALENT score. It should be noted that 20.0 is the highest and -5.0 is the lowest NLS equivalent. These limits correspond to the highest and lowest scores attainable on the NLS test.

Inspection of a scatter diagram of the NLS and TALENT Reading scores for the 50 students who were dropped from the equating sample revealed that 13 earned TALENT scores that were unusually high relative to their NLS scores and that 37 earned NLS scores that were unusually high relative to their TALENT scores. The mean and standard deviation of NLS and TALENT scores for the two groups and for all 50 students were as follows:

	NLS			TALENT		
Group	Mean	SD	N	Mean	SD	N
TALENT Higher	3.8	3.1	13	36.5	7.2	13
NLS Higher.	9.9	3.7	37	15.7	5.9	37
Total	8.3	4.5	50	21.1	11.1	50

As noted earlier, the students in the basic equating sample who took NLS first had a mean of 11.71 on that test and a standard deviation of 4.80, and the students who took TALENT first had a mean of 34.5 on that test and a standard deviation of 9.9. Thus, the excluded cases earned mean scores near the total group mean on the test for which their score was relatively high. When students were classified on the basis of the order in which they took the tests, it turned out that 28 of the 924 students who took TALENT first were excluded and 22 of the 783 students who took NLS first were excluded by the data cleaning analysis.

The effect of modifying the TALENT scores by adding random deviates can be observed by comparing the SAT-verbal means for members of the equating sample who had SAT scores stratified on the basis of each TALENT score with corresponding means for students stratified on the basis of their NLS score. The strata are defined in terms of relative standing among high school seniors in the NLS (1972) sample.

Table A3. Equivalent NLS Formula Score Corresponding to Each Modified TALENT Score

Modified TALENT score	Equivalent NLS score	Modified TALENT score	Equivalent NLS score
55	20.0	24	6.1
		23	5.7
54	20.0	22	5.2
53	20.0	21	4.9
52	20.0	20	4.6
51	20.0		
50	19.8	19	4.2
		18	3.8
49	19.4	17	3.1
48	18.9	16	2.4
47	18.2	15	1.9
46	17.6		
45	17.1	14	1.5
		13	0.9
44	16.4	12	0.4
43	15.8	11	-0.2
42	15.2	10	-1.0
41	14.6		
40	14.1	9	-1.6
		8	-1.9
39	13.5	7	-2.6
38	12.9	6	-2.9
37	12.4	5	-3.5
36	11.8	4	-4.0
35	11.2	3	-4.6
		2	-5.0
34	10.7	1	-5.0
33	10.2	(-7)-0	-5.0
32	9.7		
31	9.2		
30	8.8		
29	8.4		
28	7.9		
27	7.4		
26	6.9		
25	6.4		

Results of Subtracting SAT-Verbal Mean Based on NLS Stratification from SAT-Verbal Mean Based on Stratification by:

Stratum	Original TALENT score	Modified TALENT score
Highest tenth	+14	- 5
Second tenth	- 3	- 4
Second fifth	- 7	+ 9
Third fifth	-13	- 8
Fourth fifth	-11	+11
Lowest fifth	- 7	- 7

Because all members of the equating sample were high school seniors in 1976, the question of scale drift in SAT scores does not complicate comparisons of SAT-verbal means when students are stratified on the two reading tests, as is true in the main study. In preparing the table, each student was classified on the basis of his or her reading test scores into the appropriate ability stratum defined by 1972 high school seniors. Mean SAT-verbal scores were calculated for each subgroup so defined. Then the SAT-verbal mean for students assigned to a particular stratum using NLS scores was subtracted from the corresponding mean for students classified using TALENT scores. For example, students classified in the top stratum on original TALENT scores had an SAT-verbal mean of 566; the mean for the corresponding NLS group was 552. When modified TALENT scores were used, the SAT-verbal mean was 547. Although the pattern of results is obscured by sampling fluctuations, it appears that when students are assigned on the basis of the original TALENT scores, students in the top stratum earn higher SAT-verbal scores than students assigned to the top stratum on the basis of NLS scores. For the lower ability strata, the difference is in the opposite direction and is fairly substantial for the third and fourth fifths. A systematic difference is to be expected because the original TALENT scores are more highly correlated with SAT-verbal scores than are NLS scores. On the other hand, there seems to be no observable trend in the differences when the TALENT stratification is based on modified scores. These findings support the decision to modify the TALENT scores to make them more comparable to NLS scores.

Estimation of Equating Error

The method used for evaluating the sampling error of equating is based on the method of balanced repeated replication described by McCarthy (1969), and by Kish and Frankel (1970, 1974). For this purpose, schools were classified on the basis of stratum and testing order, making 22 groups in all, and the number of schools in each group was determined. Next, it was necessary to define two sets of subgroups, balanced as closely as possible. It was decided to include all schools in the analysis, and to allocate schools to the two sets of subgroups so that the number of schools in each set would be balanced as closely as possible with respect to group, stratum, and testing order. When the allocation process was completed, each set included 44 schools, and the number of schools assigned to one set did not exceed the number of schools assigned to the other set by more than one for any of the 22 groups, for any of the 11 strata, or either of the two testing orders. When necessary, the extra school was allocated at random to one set or the other. After the number of schools to be included in each of the 22 pairs of subgroups was determined, the assignment of schools to one or the other set was determined using random numbers. To take account of differences from school to school in the number of students in the equating sample and differences in the number of schools assigned to each of the 22 pairs of subgroups, weights were calculated so that the weighted frequencies for each subgroup would be equal. Finally, the weights for the 11 strata were adjusted to take account of the differences in the 1970 high school enrollments.

If one or the other member of a subgroup pair is chosen from each of the 22 groups, a very large number of different combinations of half-samples can be defined. For greatest efficiency in using the data, a systematic balanced design for the definition of half-samples is desirable. The necessary design principles have been developed by Plackett and Burman (1943-1945). The design appropriate to the present study is shown in Table 4A. In Table 4A, each subgroup for the TALENT-NLS testing order has a corresponding subgroup for the NLS-TALENT testing order. (Odd-numbered subgroups may be considered to have a minus sign, and even-numbered subgroups may be considered to have a plus sign, in Burman and Plackett's notation.) It will be noted that each of the 12 half-samples includes one of the two subgroups from each stratum and each testing order. The half-samples so constituted were then used to perform an equipercentile equating. This process resulted in 12 equipercentile lines of relation between NLS and TALENT Reading scores.

In order to determine the standard error of equating for a particular TALENT score, the deviations of the 12 equating results for that score from the value obtained by the equating based on the total equating were calculated. The standard error of equating is then equal to

$$\sqrt{\frac{\Sigma d^2}{12}}$$

in which "d" stands for deviation. Two decimal places were used in calculating the deviations.

The following table shows the NLS equivalent and the standard error of equating (in NLS scale units) for selected points in the TALENT scale:

Modified TALENT score	NLS equivalent	Standard error of equating
48	18.9	.15
44	16.4	.17
40	14.1	.11
36	11.8	.13

(Continued)

(Continued)

Modified TALENT score	NLS equivalent	Standard error of equating
32	9.7	.12
28	7.9	.17
24	6.1	.30
20	4.6	.32
16	2.4	.28
12	0.4	.21

For NLS scores from 7.9 through 18.9, the standard errors of equating are below .2, or about 4 percent of the standard deviation of NLS scores for high school seniors. The fact that standard errors for lower scores are larger is probably attributable to the fact, discussed earlier, that the equating sample earned relatively high scores both on TALENT and NLS.

Although Kish and Frankel (1974) concluded, on the basis of extensive empirical studies, that balanced repeated replication designs provide the best available method for evaluating sampling error for complex statistics, it should be noted that an approximation is introduced in the results by the fact that the paired subsamples were not exactly balanced with respect to the number of schools and students.

Table A4. Assignment of Subgroups to Half-Samples

Half-sample											
1	2	3	4	5	6	7	8	9	10	11	12
Subgroup											
2	2	1	2	2	2	1	1	1	2	1	1
3	4	4	3	4	4	4	3	3	3	4	3
6	5	6	6	5	6	6	6	5	5	5	5
7	8	7	8	8	7	8	8	8	7	7	7
9	9	10	9	10	10	9	10	10	10	9	9
11	11	11	12	11	12	12	11	12	12	12	11
14	13	13	13	14	13	14	14	13	14	14	13
16	16	15	15	15	16	15	16	16	15	16	15
18	18	18	17	17	17	18	17	18	18	17	17
19	20	20	20	19	19	19	20	19	20	20	19
22	21	22	22	22	21	21	21	22	21	22	21

APPENDIX B. DEFINITION OF COLLEGE ATTENDANCE

As mentioned earlier, defining college attendance posed several problems, the resolution of which resulted in attendance estimates that differ appreciably from other published estimates. The estimates we report are not necessarily incorrect; it is important, however, that our method of arriving at them be described in detail for the purposes of comparisons with other research results.

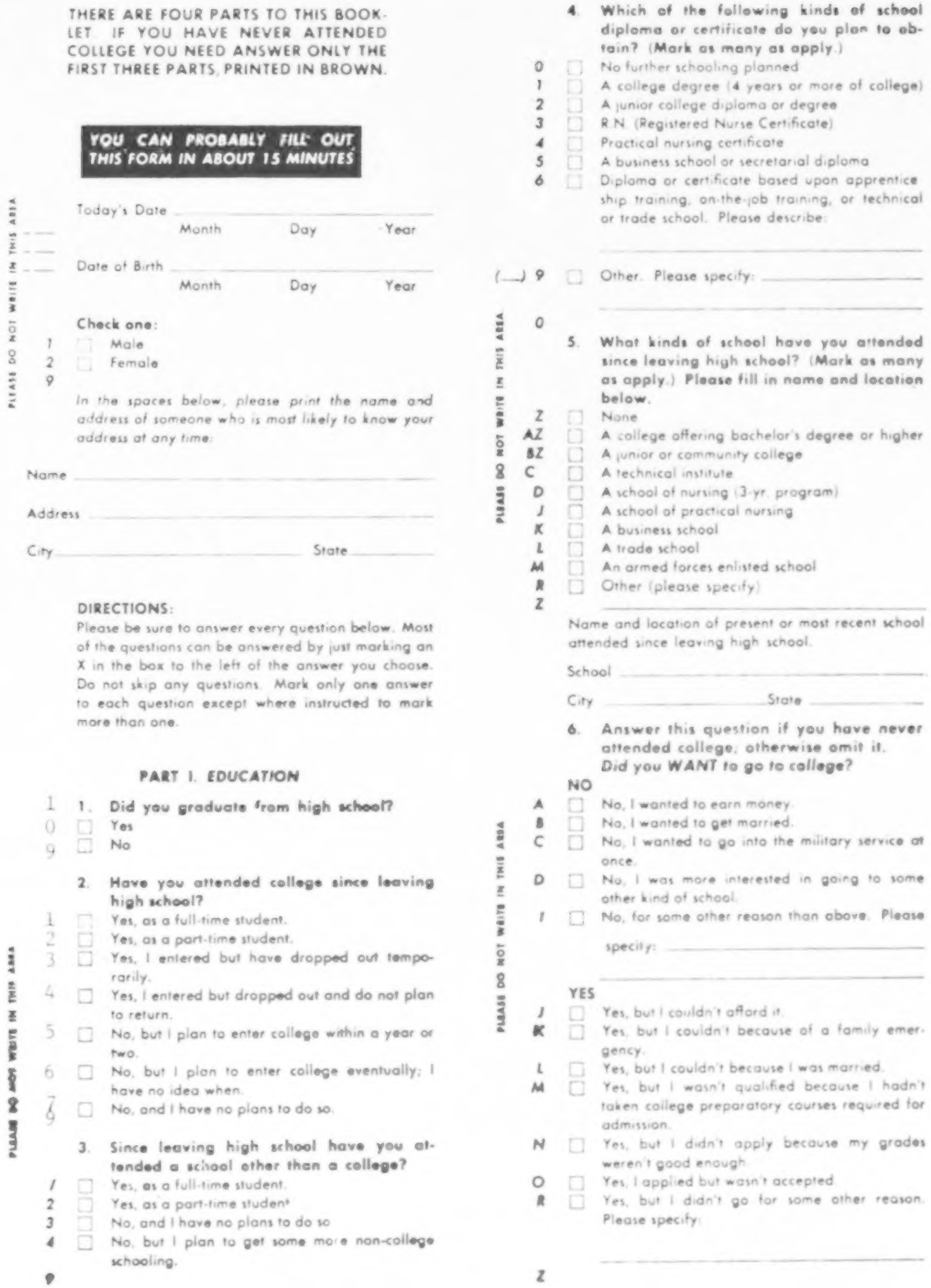
Project TALENT

For Project TALENT, college attendance was defined by means of items 2 and 5 of the first follow-up mail questionnaire, as shown in Figure B1. Only students checking boxes AZ (college offering bachelor's degree) or BZ (a junior or community college) of item 5 were considered college attendees, and then only if they did not check box 2 of item 2 (part-time student). In other words, only students who attended a two- or four-year college full time at some time during the first year following high school graduation were defined as college attendees. They may have dropped out after attending only briefly. Figure B2 is a flow diagram for the complete categorizing.

NLS

In the first follow-up questionnaire of the NLS, no less than four questionnaire items are required to categorize the students with respect to college attendance, and the respondents are directed from one item to the next by instructions that many respondents found difficult to understand, judging from the relatively large number of either omitted items or incompatible responses (Personal communication from NLS project staff). Accordingly, the project staff at Research Triangle Institute created a number of new activity state variables by pooling information from all relevant items. In addition, by examining individual questionnaires, the project staff was able to resolve discrepancies and to impute some of the missing data through logical inference. The new activity state variables were added to the NLS data release file, which includes the first follow-up data, and additional variables were later added to the file that includes the second follow-up data. For the present study, the authors used Activity State Variables 2 and 3 from the file, which include the first follow-up data (the only file available when the analysis was undertaken). These state variables were supplemented by examination of first follow-up questionnaire items 23 and 29A, in accordance with the flow diagram shown in Figure B4. This procedure produced a set of categories that to the best of our knowledge is maximally comparable to the Project TALENT categories. There are, however, some differences. For example, as shown in Figure B3, the NLS questionnaire asks about the student's educational status in the month of October 1972. Students who postponed the start of their college career until later in the year would not be included, although they would have been classified as college attendees by Project TALENT. In

Figure B1. Page 2 of Project TALENT First Follow-up Mail Questionnaire



THERE ARE FOUR PARTS TO THIS BOOKLET. IF YOU HAVE NEVER ATTENDED COLLEGE YOU NEED ANSWER ONLY THE FIRST THREE PARTS, PRINTED IN BROWN.

YOU CAN PROBABLY FILL OUT THIS FORM IN ABOUT 15 MINUTES

PLEASE DO NOT WRITE IN THIS AREA

Today's Date ____________________
Month Day Year

Date of Birth ____________________
Month Day Year

Check one:
1 ☐ Male
2 ☐ Female
9

In the spaces below, please print the name and address of someone who is most likely to know your address at any time:

Name ____________________
Address ____________________
City ____________________ State __________

DIRECTIONS:

Please be sure to answer every question below. Most of the questions can be answered by just marking an X in the box to the left of the answer you choose. Do not skip any questions. Mark only one answer to each question except where instructed to mark more than one.

PART I. *EDUCATION*

1 **1. Did you graduate from high school?**
0 ☐ Yes
9 ☐ No

2. Have you attended college since leaving high school?
1 ☐ Yes, as a full-time student.
2 ☐ Yes, as a part-time student.
3 ☐ Yes, I entered but have dropped out temporarily.
4 ☐ Yes, I entered but dropped out and do not plan to return.
5 ☐ No, but I plan to enter college within a year or two.
6 ☐ No, but I plan to enter college eventually; I have no idea when.
7 9 ☐ No, and I have no plans to do so.

PLEASE DO NOT WRITE IN THIS AREA

3. Since leaving high school have you attended a school other than a college?
1 ☐ Yes, as a full-time student.
2 ☐ Yes, as a part-time student
3 ☐ No, and I have no plans to do so
4 ☐ No, but I plan to get some more non-college schooling.
9

4. Which of the following kinds of school diploma or certificate do you plan to obtain? (Mark as many as apply.)
0 ☐ No further schooling planned
1 ☐ A college degree (4 years or more of college)
2 ☐ A junior college diploma or degree
3 ☐ R.N. (Registered Nurse Certificate)
4 ☐ Practical nursing certificate
5 ☐ A business school or secretarial diploma
6 ☐ Diploma or certificate based upon apprenticeship training, on-the-job training, or technical or trade school. Please describe: ____________________
(__) 9 ☐ Other. Please specify: ____________________
____
0

PLEASE DO NOT WRITE IN THIS AREA

5. What kinds of school have you attended since leaving high school? (Mark as many as apply.) Please fill in name and location below.
Z ☐ None
AZ ☐ A college offering bachelor's degree or higher
BZ ☐ A junior or community college
C ☐ A technical institute
D ☐ A school of nursing (3-yr. program)
J ☐ A school of practical nursing
K ☐ A business school
L ☐ A trade school
M ☐ An armed forces enlisted school
R ☐ Other (please specify)
Z ____________________

Name and location of present or most recent school attended since leaving high school.

School ____________________
City ____________________ State __________

6. Answer this question if you have *never* attended college, otherwise omit it.
Did you WANT to go to college?

NO
A ☐ No, I wanted to earn money.
B ☐ No, I wanted to get married.
C ☐ No, I wanted to go into the military service at once.
D ☐ No, I was more interested in going to some other kind of school.
I ☐ No, for some other reason than above. Please specify: ____________________
____

PLEASE DO NOT WRITE IN THIS AREA

YES
J ☐ Yes, but I couldn't afford it.
K ☐ Yes, but I couldn't because of a family emergency.
L ☐ Yes, but I couldn't because I was married.
M ☐ Yes, but I wasn't qualified because I hadn't taken college preparatory courses required for admission.
N ☐ Yes, but I didn't apply because my grades weren't good enough.
O ☐ Yes, I applied but wasn't accepted.
R ☐ Yes, but I didn't go for some other reason. Please specify:
____
Z

Figure B2. Flow Diagram for Categorizing Project TALENT Students

College Attendance: TALENT

Final Groups:
1. 4-year college, full-time
2. 2-year college, full-time
3. Vocational or other
4. No full-time school
5. Unclassified

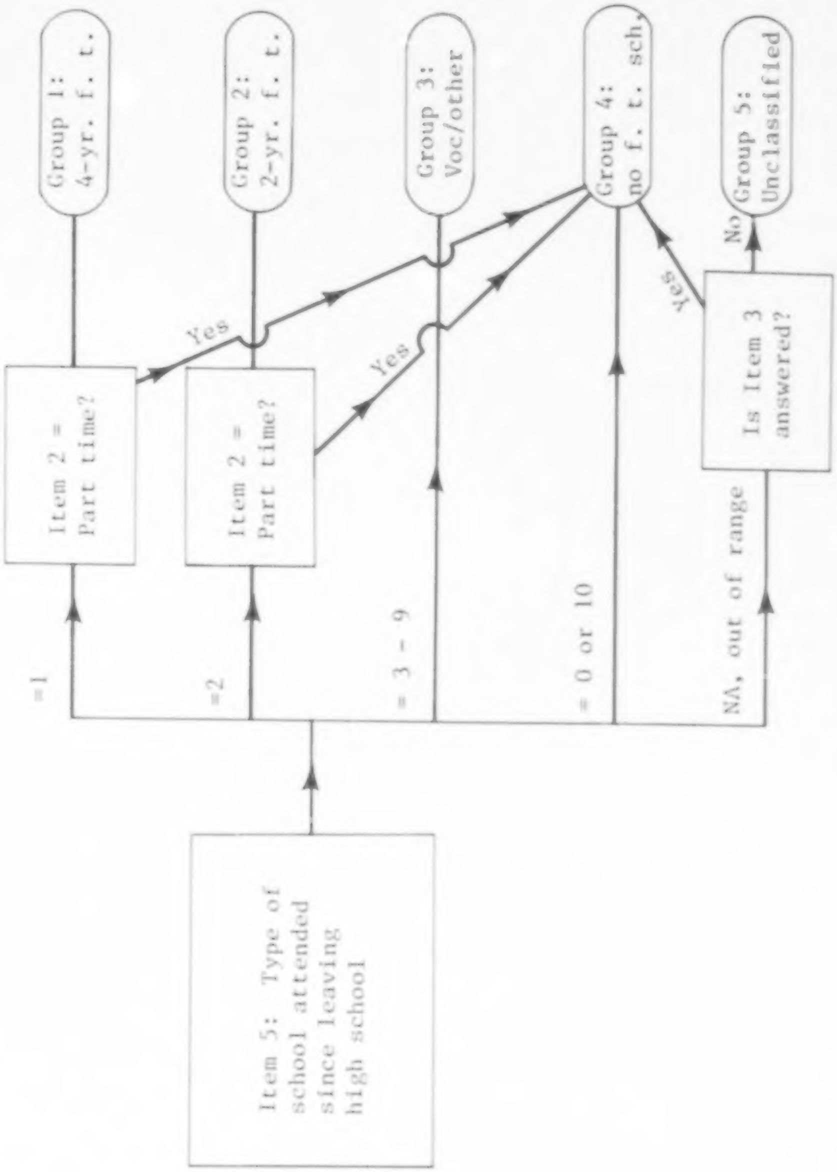

Item 5: Type of school attended since leaving high school
=1
=2
= 3 - 9
= 0 or 10
NA, out of range
Item 2 = Part time?
Item 2 = Part time?
Is Item 3 answered?
Yes
Yes
Yes
No
Group 1: 4-yr. f. t.
Group 2: 2-yr. f. t.
Group 3: Voc/other
Group 4: no f. t. sch.
Group 5: Unclassified

Figure B3. Relevant Items from NLS First Follow-up Mail Questionnaire

23. **Since leaving high school, have you attended any school like a college or university, service academy, business school, trade school, technical institute, vocational school, community college, and so forth?**

Yes	1 — *(SKIP to q. 25)*
No	2

SCHOOL ATTENDANCE IN OCTOBER 1973

25. **Were you taking classes or courses at any school during the first week of October 1973?**

No	1 — *(SKIP to q. 29a, page 9)* →
Yes	2

26a. **What is the exact name and location of the school you were attending in the first week of October 1973? (Please print and do not abbreviate.)**

School Name ____________________

City ____________________ State __________

26b. **What kind of school is this?** **(Circle one.)**

Vocational, trade, business or other career training school	1
Junior or community college (two year)	2
Four year college or university	3
Other (please describe ____________________)	4

27a. **When did you first attend this school?** __________ **(month)** __________ **(year)**

27b. **During the first week of October 1973, were you classified by this school as a full-time student?**

Yes	1
No	2

SCHOOL ATTENDANCE IN OCTOBER 1972

29a. **Now please think back a year to the Fall of 1972. Were you taking classes or courses at any school during the month of October 1972?**

Yes	1 — *(SKIP to q. 30)*
No	2

29b. **Here are some reasons others have given for NOT continuing their formal education right after leaving high school. Which of these reasons apply to you?**

30. **Was the school you attended in October 1972 the same school you attended in the first week of October 1973?** **(Circle one.)**

Yes	1 — *(SKIP to q. 33b)* *Next page* —
No, not enrolled in October 1973	2 — *(SKIP to q. 32a)*
No, enrolled in different school	3

32a. **What is the exact name and location of the school you were attending in the month of October 1972? (Please print and do not abbreviate.)**

School Name

City State

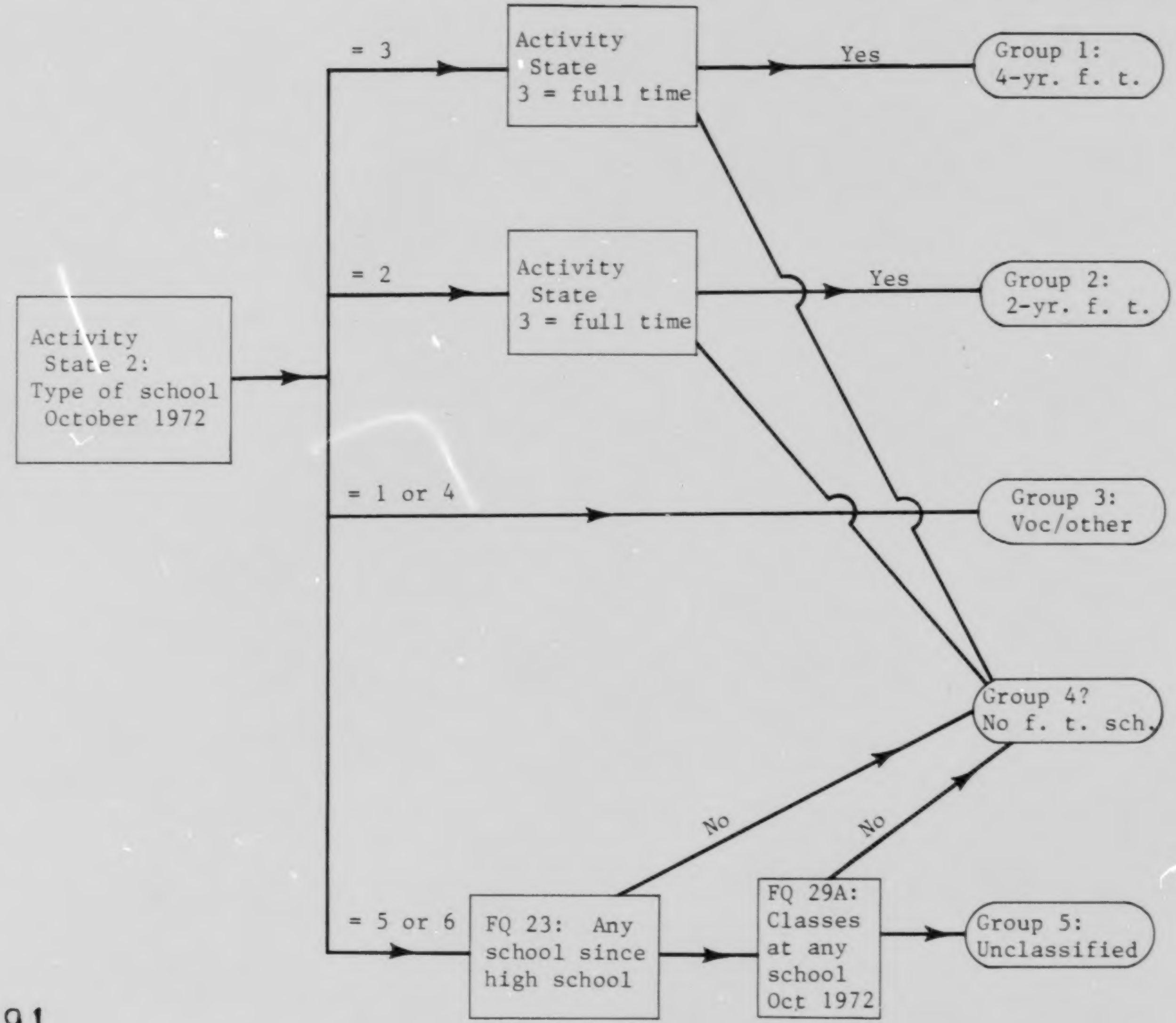
32b. **What kind of school is this?**

(Circle one.)

Vocational, trade, business or other career training school	1
Junior or community college (two year)	2
Four year college or university	3
Other (please describe ____________________)	4

Figure B4. Flow Diagram for Categorizing NLS Subjects

College Attendance: NLS


Activity State 2: Type of school October 1972
= 3
= 2
= 1 or 4
= 5 or 6
Activity State 3 = full time
Activity State 3 = full time
Yes
Yes
Group 1: 4-yr. f. t.
Group 2: 2-yr. f. t.
Group 3: Voc/other
Group 4? No f. t. sch.
FQ 23: Any school since high school
No
FQ 29A: Classes at any school Oct 1972
No
Group 5: Unclassified

191

192

general, the comparability was judged to be quite adequate for present purposes. The reader should keep in mind, however, that both surveys define college attendees in a restrictive way, namely, as students who attended a two- or four-year college full time in the year immediately following graduation.

APPENDIX C. IDENTIFICATION OF INCONSISTENT DATA CASES

Talent Sample

Early in our analyses of the 1960 Project TALENT data, we discovered an anomaly. Many students who scored very low or zero on the TALENT Reading Comprehension Test received average or above average SAT-verbal scores. In all, 54 students received reading scores in the 1-4 range, while 152 had a score of 0. Further investigation revealed that test scores were reported for Project TALENT if the test was administered, whether or not it was attempted by the student. We could not judge whether a score was valid by looking at the number of items attempted because these data were not available. (No such problem existed in the 1972 National Longitudinal Study data; in the construction of its data base, reading scores were reported only if at least one item was answered. In fact, examination of the NLS data revealed neither the cluster of low scores nor the inconsistency with the verbal sections of the SAT reported above.)

We attempted to identify and reject Project TALENT reading comprehension scores that were implausible when compared with other information about the student. Items known to be highly correlated with reading ability were chosen as predictors and are listed below under "Background Variables." Two types of comparisons, using regression analysis, were made. One used 14 criterion-scaled background and self-reported reading-related background items; the second used the other eight English tests in the test battery.

For the criterion-scaling method, we obtained the mean reading comprehension score for students choosing each option of each general background item. We also computed means for nonrespondents on each item. The actual reading comprehension score was regressed on the vector of 14 item/option means, and the regression weights were used to obtain a predicted reading score. If the difference between the actual and predicted reading scores was greater than 2.5 times the standard error, the test score was considered invalid. This eliminated 327 students of whom 35 had SAT scores.

Similarly, the reading comprehension score was regressed on the other eight English tests, a predicted score was computed, and the test rejected if a difference of more than 2.5 standard errors was found. If missing data were encountered on any of the eight tests, the same procedure was applied using each one of the first three tests, in turn. In this way, 365 students were eliminated, of whom 42 had SAT scores.

Because of considerable overlap in rejections made by the two approaches, the total number of reading test scores rejected was only 507. Of these subjects, 52 had SAT scores. All 152 of the zero scores were found to be inconsistent with the other data, thus confirming the hypothesis that a zero score really indicated a test not taken.

Background Variables

Item

56. How many books have you read in the past 12 months?
59. How many science books have you read in the past 12 months?
66. I have a difficult time expressing myself in written reports, examinations, and assignments.
73. Failure to pay attention in class has caused my marks to be lowered.
83. Slow reading holds me back in my schoolwork.
90. I read material over and over again without really understanding what I have read.
97. Number of hours studying per week.
110. My grades in English courses have been ____.
173. Estimate of your family's total income for last year.
176. How many books are in your home?
221. What is the total number of living children in your family?
237. Which one of the following best describes the college you expect to attend?
304. What is the greatest amount of education you expect to have during your life?
338. How much education are most of your friends planning to obtain?

English Tests

Vocabulary r-172
Vocabulary r-102
English
Spelling
Capitalization
Punctuation
English Usage
Effective Expression

Equating Sample

Anomalies similar to those found in the 1960 Project TALENT data also appeared in the data for the 1976 equating sample. Some students scored very high on one of the two reading tests and very low on the other. We attempted to identify and reject those students.

We performed two types of regression analyses. The first consisted of two robust regressions (Beaton and Tukey, 1974): Project TALENT Reading Test on the National Longitudinal Study Reading Test, and the NLS Reading Test on the Project TALENT Reading Test. We noted all students whose scores received weights of zero in one or both robust regressions. There were 50 such students.

The results of the robust regression were confirmed by repeating the same two regressions using least squares regression techniques. We noted all students whose scores deviated by more than 2.5 standard errors from one or both regression lines. There were 42 such cases, 40 of whom had been discovered by the robust regression.

We excluded from further analysis the 50 students identified by the robust regression analysis.

APPENDIX D. DESCRIPTION OF PARTITIONING ANALYSIS

In studying the decline in SAT scores, we asked ourselves these questions: What part of the decline is attributable to changes in the ability of high school seniors? What part to changes in the types of students taking the SAT? What part to the drift in the SAT scale itself?

Our general approach was to partition frequency distributions. The 1960 and 1972 high school seniors were classified into levels according to their reading test scores. The SAT takers were known nonrandom subsets of these frequency distributions. The difference between the frequency distributions of SAT takers in 1960 and 1972 was partitioned into sections from which answers to the above questions could be derived. All calculations were done in the metric of the 1960 SAT with the differences between 1972 and 1960 SAT-verbal averages (given equal reading scores) considered as drift in the SAT-verbal scale.

The (equated) reading scores have 20 items; thus there are 21 possible scores for the number of items right on the test and over 80 possible formula scores. We first grouped scores on the test into m intervals for calculations of frequency distributions. Presumably, m is small enough to be manageable but large enough to avoid serious loss of precision. Given this grouping of reading scores, we can present the following definitions:

N_1--the number of high school seniors in 1960
N_2--the number of high school seniors in 1971
N_{1A}--the number of SAT takers in 1960
N_{2A}--the number of SAT takers in 1972
F_1--an m^{th} order column vector containing the number of students at each level of reading ability in 1960. This is a frequency distribution of reading ability for the class of 1960.
F_{1A}--an m^{th} order column vector containing the number of students at each level of reading ability in 1960 who took the SAT.
F_2--an m^{th} order column vector containing the number of students at each level of reading ability in 1972. This is a frequency distribution of reading ability for the class of 1972.
F_{2A}--an m^{th} order column vector containing the number of students at each level of reading ability in 1972 who took the SAT.
P_1--an m^{th} order column vector containing the proportion of students at each level of reading ability who took the SAT in 1960.
P_2--an m^{th} order column vector containing the proportion of students at each level of reading ability who took the SAT in 1972.
X_1--an m^{th} order column vector containing the average SAT-verbal score for each level of reading ability in 1960.
X_2--an m^{th} order column vector containing the average SAT-verbal score for each level of reading ability in 1972.

The frequency distribution of reading scores is, therefore, represented as a column vector of length m. F_1 and F_2 represent the distributions of high school seniors in 1960 and 1972, respectively, whereas F_{1A} and F_{2A} represent the subsets of F_1 and F_2 who took the SAT. We are now prepared to manipulate these distributions.

The average SAT-verbal scores can be represented using this notation as

$$\overline{X}_{1A} = N_{1A}^{-1} F_{1A}'X_1$$

in 1960 and as

$$\overline{X}_{2A} = N_{1A}^{-1} F_{2A}'X_2$$

in 1972. Since the SAT-verbal means were computed from these same persons, these formulas are essentially the computational procedures for grouped data and are exact, not approximations. We are interested in partitioning the difference between these means into sections associated with the high school populations, SAT-takers, and the drift in the SAT scale.

According to the above definitions, the change in the distribution of reading ability for high school students is

$$\Delta F_1 = F_2 - F_1 ,$$

and the change in the proportion of students at each ability level taking the SAT is

$$\Delta P_1 = P_2 - P_1 .$$

The change in the average SAT-verbal score for persons of a given reading ability is

$$\Delta X_1 = X_2 - X_1 .$$

Before proceeding, it is important that we consider further the mean SAT score vectors, X_1 and X_2. The corresponding elements in X_1 and X_2 represent persons of nearly equal reading ability as defined by the reading tests. If the reading tests were strictly parallel to the SAT-verbal, we would expect the corresponding means to be nearly identical, i.e., $\Delta X_1 \cong 0$. The difference between these vectors is not zero because of a number of factors: sampling error, equating error, classification error, and so forth. We believe that the effect of these errors is too small to explain the observed differences between the means. We attribute most of the differences to scale drift, the slow upward growth of the SAT scores over 12 years.

What is important here is the assumption that persons with similar reading test scores are expected to have the same SAT-verbal scores regardless of whether they are members of the class of 1960 or 1972. To keep the SAT scale in a common metric over these two years, we will apply the vector X_1 to the frequency distribution F_{2A} in order to develop an SAT-verbal mean adjusted for scale drift. That mean is

$$\overline{X}_{2A'} = N_{2A}^{-1} F_{2A}'X_1$$

The frequency distribution of SAT takers in 1960 can be written as a product of the frequency distribution of high school students and the fraction taking the SAT; that is,

$$F_{1A} = F_1 \otimes P_1$$

and in 1972 as

$$F_{2A} = F_2 \otimes P_2 \ ,$$

where $\otimes$ indicates element-by-element vector multiplication. An m^{th} order unit vector 1 is useful for calculating the number of objects in a frequency distribution: Consider

$$N_{1A} = F'_{1A}1$$

and

$$N_{2A} = F'_{2A}\ 1\ .$$

Using $\Delta F_1 = F_2 - F_1$ and $\Delta P_1 = P_2 - P_1$, the frequency distribution F_{2A} can be partitioned

$$\begin{aligned} F_{2A} &= (F_2 \otimes P_2) = (F_1 + \Delta F_1) \otimes (P_1 + \Delta P_1) \\ &= (F_1 \otimes P_1) + (F_1 \otimes \Delta P_1) + (\Delta F_1 \otimes P_1) + (\Delta F_1 \otimes \Delta P_1) \\ &= G_{11} + G_{1\Delta} + G_{\Delta 1} + G_{\Delta\Delta}\ ; \end{aligned}$$

that is, the frequency distribution of SAT takers in 1972 can be partitioned into four parts:

$G_{11} = F_1 \otimes P_1$	which is identical to F_{1A}, the 1960 frequency distribution of SAT takers.
$G_{1\Delta} = F_1 \otimes \Delta P_1$	which is the part of the frequency distribution associated with a change in the pattern of SAT taking.
$G_{\Delta 1} = \Delta F_1 \otimes P_1$	which is the part of the frequency distribution associated with the change in the high school population.
$G_{\Delta\Delta} = \Delta F_1 \otimes \Delta P_1$	which is the part of the frequency distribution associated with both change in high school population and change in the proportion taking the SAT.

G_{11} is a frequency distribution, thus must have non-negative elements. $G_{1\Delta}$, $G_{\Delta 1}$, $G_{\Delta\Delta}$ are pseudodistributions and may have negative elements. Using $N_{2A} = F'_{2A}1$, we can compute

$$\begin{aligned} N_{2A} &= (G_{11} + G_{1A} + G_{\Delta 1} + G_{\Delta\Delta})'1 \\ &= N_{11} + N_{1\Delta} + N_{\Delta 1} + N_{\Delta\Delta} \end{aligned}$$

where $N_{11} = N_{1A}$, the number of SAT takers in 1960, and $N_{1\Delta}$, $N_{\Delta\Delta}$, and $N_{\Delta\Delta}$ are pseudo N's (possibly negative) associated with the pseudodistributions. We can write the 1972 mean SAT-verbal score (adjusted for scale drift) as

$$\bar{X}_{2A'} = \frac{1}{N_{2A}} F'_{2A} X_1$$

$$= \frac{1}{N_{2A}} (G_{11} + G_{1\Delta} + G_{\Delta 1} + G_{\Delta\Delta})' X_1$$

$$= \frac{1}{N_{2A}} (N_{11}\bar{X}_{11} + N_{1\Delta}\bar{X}_{1\Delta} + N_{\Delta 1}\bar{X}_{\Delta 1} + N_{\Delta\Delta}\bar{X}_{\Delta\Delta})$$

$\bar{X}_{11} = N_{11}^{-1}G'_{11}X_1 = \bar{X}_{1A}$, the mean of the 1960 SAT takers

$\bar{X}_{1\Delta} = N_{1\Delta}^{-1}G'_{1\Delta}X_1$, the effect on the 1960 mean of the changing pattern of SAT taking

$\bar{X}_{\Delta 1} = N_{\Delta 1}^{-1}G'_{\Delta 1}X_1$, the effect on the 1960 mean of the changing high school population

and

$\bar{X}_{\Delta\Delta} = N_{\Delta\Delta}^{-1}G'_{\Delta\Delta}X_1$, the effect on the 1960 mean of the changing pattern of SAT taking in the high school population.

We can now examine the three following questions:

1. What would be the number and average score of SAT takers on the SAT-verbal if the SAT-taking pattern changed as it did but the high school population stayed the same as in 1960?

The distribution would be $F_1 \otimes P_1 + F_1 \otimes \Delta P_1 = F_1 \otimes P_2$; thus the number of SAT takers would be

$\hat{N}_{12} = (F_1 \otimes P_2)'1 = N_{11} + N_{1\Delta}$,

and their mean score would be

$$\bar{X}_{12} = \frac{1}{\hat{N}_{12}}(F_1 \otimes P_2)'X_1 = \frac{N_{11}\bar{X}_{11} + N_{1\Delta}\bar{X}_{1\Delta}}{N_{11} + N_{1\Delta}} .$$

2. What would be the number of, and average score of, SAT takers if the high school population changed as it did but the proportion at each level of reading ability taking the SAT stayed the same as in 1960?

The distribution would be $F_1 \otimes P_1 + \Delta F_1 \otimes P_1 = F_2 \otimes P_1$; thus the number of SAT takers would be

$\hat{N}_{21} = (F_2 \otimes P_1)'1 = N_{11} + N_{\Delta 1}$,

and their mean score would be

$$\bar{X}_{21} = \frac{1}{\hat{N}_{21}}(F_2 \otimes P_1)'X_1 = \frac{N_{11}\bar{X}_{11} + N_{\Delta 1}\bar{X}_{\Delta 1}}{N_{11} + N_{\Delta 1}} .$$

3. What is the average scale drift?
The average score adjusted for scale drift is

$$\overline{X}_{2A'} = \frac{1}{N_{2A}} F'_{2A}X_1 ;$$

thus, the average effect of scale drift is the difference between the adjusted and unadjusted means, which is

$$\overline{X}_{2A} - \overline{X}_{2A'} = \frac{1}{N_{2A}} F_{2A}X_2 - \frac{1}{N_{2A}} F'_{2A}X_1$$

$$= \frac{1}{N_{2A}} F'_{2A} (X_2 - X_1) = \frac{1}{N_{2A}} F'_{2A} \Delta X_1 .$$

Numerical Calculations

We divided the samples of high school students in 1960 and 1972 into six strata according to reading ability, and students for whom reading test scores were not available were placed in a seventh stratum. The strata of reading scores were so defined that the top stratum contained the top 10 percent of the 1972 students for whom scores were available, the second stratum contained the next 10 percent and the lower strata contained 20 percent each. Having set the cutoff points from the 1972 sample, we allocated the 1960 sample into strata using the same cutoff points; thus, the number in each stratum is not constrained.

The basic data, contained in columns (1) through (8) of Table D1 are taken from Tables 1 and 3 of the text. The columns are

(1) = F_1 the frequency distribution of all high school students in 1960

(2) = F_{1A} the frequency distribution of SAT takers in 1960

(3) = F_2 the frequency distribution of all high school students in 1972

(4) = F_{2A} the frequency distribution of SAT takers in 1972

(5) = P_1 the proportion at each level of reading ability to take the SAT in 1960 (P_1 = (2) ÷ (1).)

(6) = P_2 the proportion at each level of reading ability to take the SAT in 1972 (P_2 = (4) ÷ (3).)

(7) = X_1 the mean SAT-verbal score for SAT takers in 1960

(8) = X_2 the mean SAT-verbal score for SAT takers in 1972.

The remaining columns are computed from (1) through (8).

(9) = (3) - (1) = ΔF_1

(10) = (6) - (5) = ΔP_1

and

$(11) = (1) \otimes (5) = G_{11} = F_{1A}$
$(12) = (1) \otimes (10) = G_{1\Delta}$
$(13) = (9) \otimes (5) = G_{\Delta 1}$
$(14) = (9) \otimes (10) = G_{\Delta\Delta}$.

As a numerical check,

$(4) = (11) + (12) + (13) + (14)$.

As noted above $(11) = G_{11}$ is a frequency distribution. The other G vectors, $G_{1\Delta}$, $G_{\Delta 1}$, $G_{\Delta\Delta}$, are the gains (or losses) in particular strata of the frequency distribution in moving from F_{1A} to F_{2A}. The effect of these partitions on the number of students is calculated by adding the frequencies in the columns, the results of which are shown at the base of those columns.

The average value of the SAT score without scale drift is derived by calculating the sum of products of these frequencies and the vector (7) $= X_1$. The effects on the means are shown at the base of the corresponding vectors.

The calculation for deriving means of hypothetical groups is given at the bottom of the table.

Note that the vectors G_{11}, $G_{1\Delta}$, $G_{\Delta 1}$, and $G_{\Delta\Delta}$ are computed for expository and interpretive purposes. The estimated populations can be computed directly by

1. SAT-taking change only

$N_{12} = (F_1 \otimes P_2)'1$

and

$\overline{X}_{12} = N_{12}^{-1} (F_1 \otimes P_2)'X_1$.

2. High school population change only

$N_{21} = (F_2 \otimes P_1)'1$

$\overline{X}_{21} = N_{21}^{-1} (F_2 \otimes P_1)'X_1$.

3. Both changed, but no scale drift

$\overline{X}_{2A'} = N_{2A}^{-1} F'_{2A}X_1$.

Table D1. Partition of SAT Scores (All Frequencies in 1000's)

Reading level	(1) F_1	(2) F_{1A}	(3) F_2	(4) F_{2A}	(5) P_1	(6) P_2	(7) X_1	(8) X_2	(9) ΔF_1	(10) ΔP_1	(11) G_{11}	(12) $G_{1\Delta}$	(13) $G_{\Delta 1}$	(14) $G_{\Delta\Delta}$
Top	255	108	288	177	.424	.614	547	567	33	.190	108	48	14	6
2	216	67	290	154	.310	.531	499	504	74	.221	67	48	23	16
3	358	74	575	226	.208	.393	438	453	217	.185	74	66	45	40
4 :	337	41	576	176	.121	.305	389	404	239	.184	41	62	29	44
5	325	18	577	121	.054	.210	341	358	252	.156	18	51	14	39
Bottom. . .	284	7	576	54	.026	.094	337	311	292	.068	7	19	8	20
No test . .	89	11	133	27	.127	.207	476	454	44	.080	11	7	6	
Sum	1864	326	3015	935					1151		326	301	139	169
Average SAT-verbal.		474		442							474	433	435	401

Estimated populations	Number	Mean
1960 SAT statistics	326	474
If only SAT taking changed	326 + 301 = 627	$\frac{326(474)+301(433)}{627} = 454$
If only high school seniors changed.	326 + 139 = 465	$\frac{326(474)+139(435)}{465} = 462$
If both changed but no scale drift	935	442
Scale drift.		$\frac{1}{N_{2A}} F'_{2A}(X_2-X_1)$ +11
1972 SAT statistics with scale drift	935	453

APPENDIX E. ESTIMATED EFFECT OF SCHOOL RETENTION ON TEST SCORES

The question of effect of school retention on test scores arises from the fact that the retention rate, from fifth grade through twelfth grade, was 66.7 percent in 1960 but rose to 79.1 percent in 1972 (Digest of Education Statistics, 1975 Edition, Table 10). If we assume that the additional students "retained" in the schools tended to be of lower ability, then a sizable share of the decline in reading scores for the high school population might be explained.

To obtain the required estimates, we needed an estimate of the correlation between dropping out and reading ability. We obtained this by reanalyzing some drop-out data obtained from one of the school systems that participated in the Growth Study from 1961 to 1969. The biserial correlation is - .525.

Using this correlation and the population proportions noted above, we then estimated, by means of the standard formula for the biserial correlation, how much the mean of the test scores would be lowered by a decrease in the number of school dropouts from 1960 to 1972. This lowering would be .095 of the standard deviation (S.D.) of the school cohort.

In the Growth Study data, the S.D. of the school cohort mentioned above is 4 percent larger than the S.D. of the population of surviving twelfth graders. If we assume that the same proportion holds in the present study, we can estimate an S.D. of 5.2. The score drop attributable to the increased retention rate would, then, be .5 points. We observed a total drop for the high school senior population of .8 points. Thus, it would appear that the increased retention alone may account for approximately 60 percent of the total decline

Further research is needed to determine whether this estimate would be confirmed by a formal study of the question.

An Investigation of Item Obsolescence in the Scholastic Aptitude Test

James Braswell
Nancy Petersen
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

January 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $5 per copy, as well as a price list for all the appendixes can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

ACKNOWLEDGMENTS

The College Board and Educational Testing Service wish to express their appreciation to the following panel members who reviewed verbal and mathematical test questions.

Verbal Panel

William Dwyer
Staff Writer, Trentonian
Trenton, New Jersey

Paul Glatzer
History and Anthropology Teacher
Walt Whitman High School
Huntington Station, L.I., New York

Hugh Lane
Research Associate
Institute for Studies in Education
Washington, D.C.

Jean Malmstrom
Department of English
Western Michigan University
Kalamazoo, Michigan

Sharilyn McSwan
Language Arts Resource Teacher
Grades 7-12, San Diego Public Schools
San Diego, California

Maria Penny
Physics Teacher
Wilde Lake High School
Columbia, Maryland

Charles Suhor
Supervisor of Secondary English
New Orleans Public Schools
New Orleans, Louisiana

Mathematics Panel

Betty Hinman
Division of Mathematics
University of Houston
Houston, Texas

David Johnson
Mathematics Teacher
Nicolet High School
Milwaukee, Wisconsin

Katherine Layton
Mathematics Teacher
Beverly Hills High School
Beverly Hills, California

Willie May
Mathematics Teacher
Wendell Phillips High School
Chicago, Illinois

Eugene Nichols
Department of Mathematics Education
Florida State University
Tallahassee, Florida

William E. Slesnick
Department of Mathematics
Dartmouth College
Hanover, New Hampshire

Robert Smith
Mathematics Teacher
Lawrence High School
Lawrenceville, New Jersey

ABSTRACT

Two panels, one verbal and the other mathematical, were appointed to review and rate questions that appeared in earlier editions of the Scholastic Aptitude Test (SAT) and then in more recent editions. The elapsed time between test administration dates varied between three and 13 years. Raters were sent copies of the various tests and were asked to use a five point rating scale to indicate how the difficulty of each

question might be expected to change between the two administration dates. The raters' predictions were then compared with available item analysis data. Of particular interest were those questions for which the item analysis data showed a substantial change in performance and the raters predicted and could explain the change.

The general trend was for verbal raters to rate questions as being more difficult for recent groups while the trend for mathematical raters was to rate the questions as being somewhat easier for recent groups. Inter-rater correlations were generally low. Rater means were computed for each type of verbal and mathematical question by test. No significant differences were found in the verbal means, but on two of the mathematical tests there were significant differences in rater means due primarily to a higher rating assigned to questions classified as "miscellaneous." A similar analysis was done using the actual item analysis data. On none of the mathematical tests and on only one of the verbal tests was there a significant difference in actual student performance by type of question. The more recent group's performance on the analogies was relatively better than the earlier group's performance.

Because the mathematical questions are more closely related to specific school-based learning than are the verbal questions, one might predict that the mathematical raters' judgments would correlate higher with actual performance based on item analysis data than would those of the verbal raters. Such was the case. The verbal correlations were uniformly low for all raters and for both tests (the greatest was 0.16). For two of the four mathematical tests in which the time spread between administrations was relatively long, several of the mathematical raters' judgments correlated reasonably well with item analysis data (the greatest was 0.66). This finding suggests that curricular changes in mathematics over the past 15 years have had a differential impact on questions administered in the 1960s and then re-administered in the 1970s.

The report also contains 20 specific questions for each of verbal and mathematical that had extreme (either high or low) rater means or were identified as being outliers based on the item analysis data. No verbal questions identified at the extremes in terms of rater means were also identified as extreme outliers in the statistical analyses. Three of the 10 mathematical questions identified as extreme in terms of rater means were also identified as outliers in the statistical analyses.

INTRODUCTION

For the past several years both verbal and mathematical scores on the Scholastic Aptitude Test (SAT) have been declining. The decline is a direct result of lower performance of more recent groups of students on the various equating tests. At each administration of a new edition of the SAT, equating questions that appeared in a previous edition of the test one to six years earlier are used to place raw scores on the new operational test on the 200-800 score scale. These questions are usually placed in a separately timed nonoperational section of the test. Separate equating is done for the verbal and mathematical parts. The verbal equating tests contain 40 questions that are representative in content and difficulty of the operational verbal test. The mathematical equating tests contain 25 questions that are likewise representative of the operational mathematical test.

Because score decline is a direct result of lower average performance on equating questions that are reused over a period of several years, it is important to examine the content of representative questions used over time for possible obsolescence. Are there factors, curricular or otherwise, that would make certain questions less relevant today than they were in the past? Are there some questions that might be more relevant today?

The primary purpose of this study was to identify and examine those verbal and mathematical questions that changed in difficulty between administration dates. While it has been noted that the mean scores on equating tests have been gradually declining, it has also been observed that performance on some questions remained relatively stable. In contrast to the general trend, a few questions have actually become easier for more recent groups of students.

DESIGN OF THE STUDY

Two independent panels of raters, one for verbal and one for mathematical, were appointed. Raters were sent copies of the various tests and were asked to use a five point rating scale to indicate how the difficulty of each question might be expected to change over the given time period. In addition, the rating form provided space for the rater to comment on the probable cause of the predicted shift in difficulty. The raters' predictions were then compared with the available item analysis data for the two administration dates of each set of questions. Of particular interest were those questions in which the item analysis data showed a substantial change in performance and the raters predicted and could explain the change.

TEST FORMS

Two complete 90 question verbal operational tests and four complete 25 question mathematical equating tests were used in the study. The two verbal tests were those used in the Modu-Stern SAT scale stability study. The earlier and later administration dates are shown in the chart at the top of the next page.

Test	Number of questions	Administration dates Earlier	 Later
Verbal 1	90	May 1963	May 1976
Verbal 2	90	March 1973	March 1976
Math 1	25	May 1962	January 1971
Math 2	25	January 1965	December 1973
Math 3	25	April 1971	November 1975
Math 4	25	March 1972	April 1976

TEST REVIEW PANELS

The SAT-verbal is a general test of developed verbal ability. In order to insure that the test discriminates fairly among students, no matter what their particular field of interest, specifications are set so that the content of each test includes material from a variety of fields. Discrete questions, questions that are not related to passages, are taken from the fields of aesthetics and philosophy, human relationships, science, and the world of practical affairs. Each edition of the SAT-verbal has a narrative passage, a humanities passage, a science passage, an argumentative passage, and a social studies passage. It follows, therefore, that the verbal questions should be reviewed by people with different interests in order to insure that comments are broad and far-reaching.

The verbal review panel consisted of seven members--four generalists and three specialists. Those classified as generalists were a high school physics teacher, a high school social studies teacher, a newspaper writer, and an educational researcher. Those classified as specialists were a member of a university English department, a language arts curriculum coordinator, and a curriculum specialist in a large city school system.

Because the mathematical portion of the SAT assumes certain elementary content prerequistes in arithmetic, algebra, and geometry, it was decided that the seven member mathematical panel would consist of four high school mathematics teachers who would be aware of what is and has been taught and emphasized and three specialists at the university level who have been aware of major curricular trends in mathematics over the past 10-15 years. Two of the specialists are authors of textbooks at the high school level. All three specialists are involved in either teaching mathematics, training mathematics teachers, or conducting related research. Two of the panel members, one high school teacher and one college teacher, are currently serving on a panel appointed by the Mathematical Association of America to study the decline in the mathematical preparation of students entering college.

RATING INSTRUCTIONS

Separate detailed instructions for reviewing the test questions and test review forms were prepared for the SAT-verbal and the SAT-mathematical tests. Copies of these instructions are included in Appendix A. Each panel member was asked to consider each question in the various test booklets and to judge whether or not students would have found the question easier or harder at the second administration than at the first administration. The administration dates were given on the cover of each test booklet. In making each judgment, the raters were asked to assume that the earlier and the later student groups were equal in ability and that the change they predicted in difficulty level was due not to any change in the ability of the students, but to a change in the meaning of the question or the familiarity of the material on which it was based.

The following codes were used by each panel member to indicate his or her judgment about the change in difficulty for each question:

++ Denotes questions that should be considerably easier at the more recent administration date than at the earlier administration date.

+ Denotes questions that should be somewhat easier at the more recent administration date than at the earlier administration date.

0 Denotes questions that should be approximately the same difficulty.

- Denotes questions that should be somewhat more difficult at the more recent administration date.

-- Denotes questions that should be considerably more difficult at the more recent administration date.

A review form for coding responses was provided for each test. Space was also provided on the form for the reviewers to indicate the extent to which curricular changes may have contributed to changes in difficulty during the time between the two administration dates. This "Comments" section allowed the reviewers to indicate specific reasons (curricular or societal) for a predicted change in the difficulty level of a question. Reviewers were particularly encouraged to comment on questions coded either ++ or --.

THE DATA

There are two types of data for each test question: (1) the raters' judgments regarding the change in difficulty of the question over time, and (2) a delta value, a measure of the question's difficulty based on the fraction of candidates responding correctly to the question at each administration of the test. We will first discuss the judgments made by the raters and then compare their judgments with the observed changes in question difficulty.

In the analysis to follow, the codes ++, +, 0, -, and --, used by each panel member to indicate his or her judgment about the change in difficulty for a question, have been recoded 5, 4, 3, 2, and 1, respectively, to enable numerical analysis of the judgments. The raters' judgments for each question within each test are recorded in Tables B1-B2 (see Appendix B). With the exceptions of verbal raters one and three, each rater tended to code most questions 3 indicating that they thought most questions should have been of approximately the same difficulty at each administration. The means and standard deviations of the ratings for each panel member by test and across all tests (two for verbal and four for mathematical) are given in Table B3 (see Appendix B). The average rating for the verbal questions was 2.78 indicating that, overall, the verbal raters felt that candidates should have found the questions slightly more difficult at the more recent administrations. The mean of 2.17 for verbal rater three was considerably lower than the mean for any other verbal rater. The large standard deviation of 1.06 for verbal rater one indicates that this rater tended to use the full range of codes in making judgments about questions.

The average rating of 3.11 for the mathematical questions indicates that, overall, the mathematical raters felt that candidates should have found the mathematical questions slightly easier on the more recent administrations. Mathematics Test 2 contained more questions than the other tests that raters predicted would be relatively easier at the more recent administration (1965-1973). The rater means for this test were generally higher than the means for the other three mathematics tests. The reasons given by raters in the space provided on the rating form were uniformly related to curricular changes--changes that would in most cases benefit students who took the test in 1973. There were some individual questions with low rater means and reasons given for the predicted decline in performance also related to changes in curricular emphases. As can be seen later when individual mathematics questions are examined, there was some consistency of agreement between mean rater judgments and the actual statistical results derived from item analysis data.

The inter-rater correlations by test and across all tests are given in Table B4 (see Appendix B). These correlations excluded items that were extremely easy or extremely hard because such questions are subject to greater chance variation. This is especially true for the more difficult questions because of random guessing. The excluded questions were those with a ** in Tables B5-B10. The inter-rater correlations tend to be small for verbal questions. This is partly due to the lack of variability in the ratings assigned to the various questions. Each panel member correlates positively, indicating agreement, with the mean panel rating. However, there are a few negative correlations between panel members. Some raters tended to code most of the questions on a test 3, indicating no change in difficulty. Consequently, when these raters coded a question other than 3, they disagreed a couple of times with other panel members. These few disagreements tended to cause the small negative correlations. The average inter-rater verbal correlation was 0.12.

The mathematical inter-rater correlations are generally low to moderate. There was considerably more agreement among the mathematical raters than the verbal raters. This is probably due in part to the fact that all seven members of the mathematical panel were in that discipline. The average inter-rater mathematical correlation was 0.25.

The deltas, distance from the major axis, and the mean and standard deviation of panel ratings for each question are given in Tables B5-B10 (see Appendix B). The information in these tables is explained and discussed below.

The difficulty level of the test questions at Educational Testing Service is reported on a "delta" scale. The more difficult the question the greater the delta value. The relationship between delta and the traditional percent correct can be found in Figure B1 (see Appendix B). Except for Verbal Test 2 and Mathematics Test 4, the mean delta for the more recent administration of the tests was larger than that for the earlier administrations, indicating that the more recent groups found the test more difficult. Given that the tests were, on the average, more difficult for the more recent groups, one goal of the study was to identify those questions that were outliers and then to examine the raters' judgments about these questions.

In order to identify outliers, the deltas were plotted so that each point on the graph represents the pair of deltas observed at the two administrations of the question. The plot of these points normally forms an elongated ellipse extending from the lower left corner to the upper right corner. If the candidate groups taking the test at each administration are from similar populations, the ellipse is long and narrow indicating a high correlation. When the groups differ in mean ability level, the ellipse will be displaced above or below the 45° diagonal line (equal-difficulty line) depending on which group is the more able. The greater the discrepancy between the two groups in the way they respond to individual questions, the farther the corresponding points will lie from the major axis line of the ellipse. The perpendicular or shortest distance of the point from the major axis line is a measure of this discrepancy.

To illustrate, the delta plot for Verbal Test 1 is given in Figure B2 (see Appendix B). The horizontal axis represents the more recent administration, and the vertical axis represents the earlier administration. The major axis of the ellipse lies parallel and below the equal-difficulty line (the broken 45° diagonal line) indicating that the more recent group was less able than the earlier group. Points which fall on the major axis indicate questions which probably represent the same task for both groups. Points above the major axis indicate questions which were relatively more difficult for the earlier group and have negative perpendicular distances associated with them, whereas those below the line indicate questions which were relatively more difficult for the more recent group and have positive perpendicular distances associated with them.

Not all questions were used in the construction of the delta plots. The usefulness of a delta plot depends on the soundness of the assumption that the major axis of the ellipse adjusts for differences in ability level between the two groups and that its location is unaffected by other confounding factors such as item obsolescence or error generated by guessing. In an attempt to remove any such biases, all questions for which the mean rating was greater than 3.3 or less than 2.7 were deleted from construction of the delta plot, as such ratings might be indicative of questions that have changed in meaning. Also, all questions for which the delta values in at least one administration were greater than 15.7 or less than 6.4 were deleted. A delta of less than 6.4 corresponds to 95 percent or more of the candidates responding correctly to the question whereas a delta greater than 15.7 corresponds to 25 percent or less of the candidates responding correctly, that is, the guessing range. This last group of questions were deleted from all remaining analyses of the data.

ANALYSIS OF THE DATA

Each of the verbal questions is categorized as either sentence completion, antonym, analogy, or reading passage. Each of the mathematical questions is categorized as either arithmetic, algebra, geometry, or miscellaneous. It is of interest to determine whether (1) the raters judged particular types of questions as becoming easier or more difficult, (2) the perpendicular distances indicate that particular types of questions have become easier or more difficult, and (3) whether the two indices agree. It is also of interest to determine whether the raters or the perpendicular distances indicate more change in question difficulty as the time between administrations of the questions increases.

The one-way analysis of variance technique was used to study the above questions for each test. The one-way analysis of variance technique is a method for comparing means when there are two or more categories of interest. The hypothesis being considered is that the means for each category are identical. The test statistic used is called the F ratio. The F ratio is basically the ratio of the variance between the means for each category to the variance between observations within each category. If the hypothesis is false, that is, if the means for the categories really do differ, then the differences among the means will be large in relation to the differences among observations in each category and the resulting F ratio will be large. The larger the F ratio the smaller the probability that the means for each category differ only by chance.

Table 1 summarizes the one-way analysis of variance results for the panels' average question ratings categorized by question type. Except for Mathematics Tests 2 and 3, none of the F ratios are large enough to say that the differences in the means for the various types of questions are due to something other than chance, that is, there is not enough evidence to say that raters tend to view any particular type of question as having changed in difficulty over time. However, for Mathematics, there is a trend. The average rating for the "miscellaneous" questions is always the largest, and being greater than 3, this indicates that the raters thought these questions would be easier at the more recent administration.

The one-way analysis of variance results for the perpendicular distances categorized by question type are summarized in Table 2. Except for Verbal Test 2, the F ratios are nonsignificant. This time there is no consistent trend in the means for the mathematical questions. However, there is a trend for the verbal questions. The analogy type question appears to have been easier at the more recent administration (negative mean perpendicular distance). Thus, neither the analysis of the rater data nor the analysis of the distance data consistently identified types of questions such as analogy or algebra as having changed in meaning between test administrations.

Table 1. One-Way Analysis of Variance
Average Rating Categorized by Question Type

Verbal

Test	Summary statistics	Question type				F ratio
		Sentence completion	Antonym	Analogy	Reading passage	
1	n	16	12	17	29	
	mean	2.77	2.63	2.78	2.87	1.40
	sd	.40	.28	.24	.36	
2	n	12	13	14	33	
	mean	2.95	2.91	2.79	2.80	2.16
	sd	.22	.12	.25	.21	

Mathematics

Test	Summary statistics	Question type				F ratio
		Arithmetic	Algebra	Geometry	Misc.	
1	n	8	2	6	0	
	mean	3.02	2.86	3.05		.34
	sd	.31	.29	.16		
2	n	4	9	5	4	
	mean	3.07	3.30	3.31	4.00	4.08*
	sd	.50	.33	.21	.43	
3	n	5	7	3	2	
	mean	3.03	3.14	3.05	3.36	3.70*
	sd	.06	.11	.18	.07	
4	n	5	6	6	3	
	mean	3.08	3.07	3.07	3.19	.84
	sd	.11	.07	.14	.07	

*Significant at 5 percent level.

Table 2. One-Way Analysis of Variance
Distance from Major Axis Categorized by Question Type

Verbal

		Question type				
Test	Summary statistics	Sentence completion	Antonym	Analogy	Reading Passage	F ratio
1	n	16	12	17	29	
	mean	.01	.03	-.28	.03	1.56
	sd	.37	.57	.55	.47	
2	n	12	13	14	33	
	mean	.06	.04	.17	.04	2.75*
	sd	.24	.27	.34	.16	

*Significant at 5 percent level.

Mathematics

		Question type				
Test	Summary statistics	Arithmetic	Algebra	Geometry	Misc.	F ratio
1	n	8	2	6	0	
	mean	-.03	-.28	.09		.58
	sd	.31	.38	.46		
2	n	4	9	5	4	
	mean	.35	-.19	-.01	-.20	1.16
	sd	.42	.57	.23	.42	
3	n	5	7	3	2	
	mean	.09	-.07	.09	-.15	.56
	sd	.33	.23	.19	.14	
4	n	5	6	6	3	
	mean	.07	-.16	-.07	.33	2.24
	sd	.16	.20	.37	.09	

The time between the two administrations for the various tests differs substantially. Thirteen years elapsed between the two administrations of Verbal Test 1 whereas only three years elapsed between the two administrations of Verbal Test 2. Similarly, nine and eight years elapsed between the two administrations of Mathematics Tests 1 and 2, respectively, whereas only four years elapsed between the two administrations of Mathematics Tests 3 and 4. Perhaps, by analyzing the data by time between administrations, one can determine whether the raters thought there should be more change in the difficulty level of a question as the number of years between administration increases.

Table 3 summarizes the one-way analysis of variance results for the panels' average ratings categorized by tests. Mathematics Tests 1 and 2 and Mathematics Tests 3 and 4 were combined for each analysis as the time span between the two administrations for each set of tests was approximately the same. Neither of the F ratios is large enough to suggest that the difference in the mean ratings are due to something other than chance.

Since the analyses using the ratings and the analyses using the distances do not suggest the same sort of trends, it is of interest to see just how much agreement there is between the two measures. Table 4 summarizes the one-way analysis of variance results for perpendicular distances categorized by average rating. If the two types of data agree, then those questions with a rater mean greater than 3 should have a negative mean perpendicular distance, those questions with a rater mean of 3 should have a mean perpendicular distance of approximately zero, and those questions with a rater mean less than 3 should have a positive mean perpendicular distance. The F ratios were nonsignificant, that is, there is not enough evidence to say that the

Table 3. One-Way Analysis of Variance
Average Rating Categorized by Test

Summary statistics	Verbal		F ratio
	Test 1	Test 2	
n	74	72	
Mean	2.79	2.84	1.40
sd	.34	.22	

Summary statistics	Mathematics		F ratio
	Test 1 and 2	Test 3 and 4	
n	38	37	
Mean	3.23	3.10	2.66
sd	.44	.13	

Table 4. One-way Analysis of Variance
Distance from Major Axis Categorized by Average Rating

Verbal

Test	Summary statistics	Average rating			F ratio
		Greater than 3	Equal to 3	Less than 3	
1	n	13	13	48	
	mean	.22	-.17	-.08	2.47
	sd	.60	.60	.41	
2	n	11	17	44	
	mean	-.01	.02	-.00	.05
	sd	.17	.25	.27	

Mathematics

Test	Summary statistics	Average rating			F ratio
		Greater than 3	Equal to 3	Less than 3	
1	n	7	3	6	
	mean	-.32	.14	.26	1.48
	sd	.34	.15	.27	
2	n	17	2	3	
	mean	-.17	.67	.13	3.18
	sd	.43	.16	.56	
3	n	9	7	1	
	mean	-.06	.10	-.16	.84
	sd	.24	.28	.00	
4	n	13	5	2	
	mean	.09	-.16	-.23	2.04
	sd	.29	.10	.40	

means differ other than by chance. Also, the hypothesized trend in the means described above is not even present in each data set.

For an item to have a rater mean greater than 3, for example, it is not necessary for all raters to rate the question as either 3, 4, or 5. It is possible, then, that lack of rater consistency is clouding the results. To see if this is the case, an analysis similar to that summarized in Table 4 was done, using only those items for which the raters' judgments were consistently 3 or more, 3 or less, or equal to 3. However, again the F ratios were nonsignificant.

The correlations of rater judgments on individual questions with the distance of that question from the major axis are given in Table B4 for each test. This data is derived by correlating the judgments of each rater in Tables B1 and B2 with the corresponding distance values in Tables B5-B10. Recall that a question rated 5 is judged to be considerably easier for the more recent group while a question rated 1 is judged to be considerably more difficult. Also, recall that outliers that have negative distances associated with them are relatively more difficult for the earlier group while those with positive distances associated with them are relatively more difficult for the recent group. Therefore, a rater judgment of 5 would be consistent with an outlier distance value of -1.1 and a rater judgment of 2 would be consistent with a distance value of +0.4. Thus, a large negative correlation indicates agreement.

The verbal correlations are generally quite small and sometimes in the opposite direction from what one would expect. In fact, the correlation between the average rating and the distance from the major axis on Verbal Test 1 is 0.13 and on Verbal Test 2 is 0.01 which indicate slight disagreement between means and distances.

Some of the mathematical correlations are surprisingly high. For example, on Mathematics Test 1, the average rating and raters one and two had correlations of -0.60, -0.57, and -0.56, respectively, with the distances and on Mathematics Test 2, the average rating and raters one and five had correlations of -0.61, -0.66, and -0.44, respectively, with the distances, indicating a reasonable amount of agreement between raters' judgments and actual distance results. The correlations for Mathematics Tests 3 and 4, on the other hand, are uniformly small.

Given the rather large correlations between average rating and distance from major axis on Mathematics Tests 1 and 2, it seems somewhat surprising that the result of the analysis in Table 4 is nonsignificant. Figures B3 and B4 are scatterplots of the average rating versus the distance from the major axis for Mathematics Tests 1 and 2, respectively. In Figure B3, questions 4, 7, 13, and 14 appear to be outliers. If these four questions are deleted, the correlation decreases from -0.60 to -0.13. In Figure B4, questions 2 and 11 appear to be outliers. If these two questions are deleted, the correlation decreases from -0.44 to -0.11. Thus, the rather large correlations between average rating and distance from major axis on Mathematics Test 1 and 2 seem to be largely due to the presence of a few questions which were outliers having consistent extreme average ratings and extreme distance values.

A LOOK AT PARTICULAR QUESTIONS

This section of the report will examine those specific questions that had either high mean rankings, low mean rankings, or were identified as being outliers on the delta plots. The raters made many extremely interesting comments on individual questions. Although most of the comments were not confirmed by the delta plots and F-tests, many of the comments will be useful in future test assembly.

Tables C1-C2 in Appendix C present those five verbal questions that had the highest rater means and those five questions that had the lowest rater means. Questions with a high rater mean represent questions that the raters predicted would become relatively easier for the more recent groups while those with low rater means would become harder. Included for each question are the rater mean, the perpendicular distance, and the specific comments made by raters. For those questions that relate to a reading passage, the entire passage is shown.

Tables C3-C4 present the same type of information for the five verbal questions at the greatest negative distances from the major axis and the five questions at the greatest positive distances. The negative distances favor the more recent groups while the positive distances favor the earlier groups. No question identified in Table C1 is also identified in Table C3. This means that none of the top five questions judged by the rater means as most likely favoring the more recent groups were in the top five questions identified in the delta plot analysis. Likewise, there were no common questions identified in Tables C2 and C4. Thus, none of the top five verbal questions identified by the rater means as favoring the earlier groups were identified in the corresponding delta plot analysis. There are, however, some trends that run in the same direction. On Test 1 the passage for questions 86-90 was computer-related. Several of the raters felt that today's students have greater knowledge of computers and related vocabulary. Questions 87 and 90 had relatively high rater means and each was at a negative distance of -0.49 from the major axis, indicating these questions were in fact relatively easier for the more recent group.

It should be noted that questions at the extremes of the difficulty scale were considered in this portion of the analysis and that chance factors alone could be responsible for observed delta changes on difficult questions. Most of the extremely difficult vocabulary questions that raters generally ranked as more difficult for the recent group actually turned out to be relatively less difficult. A primary reason for this could have been the random noise generated by guessing.

Tables C5-C6 in Appendix C show those five mathematical questions that had the highest rater means and those five questions with the lowest rater means. Tables C7-C8 show those five mathematical questions that had the greatest (negative) perpendicular distances from the major axis and the five questions that had the greatest (positive) perpendicular distances. Of the five questions with the highest rater means (indicating questions that should be easier for more recent groups), two of these questions were also identified as extremes in the delta plots. These were questions 5 and 11 on Test 2. Of the five mathematical questions with the least rater means (indicating questions that should be harder for more recent groups), only one of these questions was also identified as an extreme by the delta plots. This was question 2 on Test 2.

There were two areas where the mathematical raters' opinions were supported by the delta plots. The raters stated for questions 5 and 11 on Test 2 that greater recent emphasis on real number properties, properties of zero, and abstract operations, should have made these questions easier for more recent groups. These questions were also identified in the delta plots as questions that were relatively easier for the more recent group. Question 2 on Test 2 which involves percent, was identified in the top five by both raters and delta plots as being more difficult for the recent group. The rater comments (see Table C5, Appendix C) do not suggest why the question would be more difficult, but comments by raters on other percent questions suggest that this topic is receiving less emphasis on the curriculum.

IMPLICATIONS OF THE STUDY

In order to relate raw scores on each new edition of the SAT to the established 200-800 score scale, separately timed verbal and mathematical "equating" tests are readministered in a nonoperational section of the SAT. Through the equating process, raw scores on the new test are adjusted for the estimated difference between the ability of the group that took the new test and the ability of a prior reference group. A typical equating test is used three to five times over a five year period and then retired. The equating process assumes that questions readministered represent the same psychological task at each administration. This study was conducted to determine whether there are intervening curricular or societal factors that might produce equating error.

Representative verbal and mathematical questions were rated by two independent panels and no evidence was found to suggest that identifiable curricular and societal factors are responsible for short term (three to five years) observed changes in item performance. The same was true for verbal items, whose administration dates were 13 years apart. However, for the mathematical questions, there is evidence to suggest that identifiable curricular factors may account for the differential performance on a few of the questions. In view of this evidence, the current practice of retiring equating tests after about five years seems appropriate, especially for the mathematical equating tests.

SUMMARY

Score decline is a direct result of lower average performance on equating questions that are reused over a period of time. It is possible that the lower performance might be explained by curricular or societal changes that would make certain questions less relevant today than they were in the past. The primary purpose of this study, then, was to identify those verbal and mathematical questions that have changed substantially in difficulty between administrations.

Two 90 question verbal operational tests and four 25 question mathematical equating tests were used in the study. Questions were identified as outliers using (1) judgments made by raters and (2) perpendicular distances calculated from delta-plots.

The ratings and the perpendicular distances were both analyzed by type of question. Analysis of the ratings suggested that the raters

thought that mathematical "miscellaneous" questions would have been easier at the more recent administrations. Analysis of the perpendicular distances indicated that verbal analogy questions tended to be easier at the more recent administrations. The conclusions from the two analyses by type of question do not tend to support each other. The ratings were also analyzed by elapsed time between test administrations. The analysis failed to suggest that differences in rater means for the two time periods were due to something other than change.

Analysis of the perpendicular distances classified by average rating showed that raters were, on the whole, unable to predict the observed change in difficulty level of a question between administrations. This result seemed somewhat incongruous with the high correlations between average rating and distance from major axis for Mathematics Tests 1 and 2. However, examination of the scatter plots showed that the high correlations of average rating with distance from major axis for Mathematics Tests 1 and 2 were largely due to a few questions that were outliers having consistent extreme average ratings and extreme distance values.

Also examined were each of the five verbal and mathematical questions with (1) the highest mean rating, (2) the lowest mean rating, (3) the greatest negative perpendicular distance, and (4) the greatest positive perpendicular distance. None of the verbal questions and only three of the mathematical questions were identified as outliers in the same direction by both sets of data.

This study was conducted to see whether raters might be able to predict and explain changes in the difficulty level of questions over time. For the most part, the changes predicted by raters were not substantiated by the statistical analysis. Therefore, while the relative difficulty of some questions changed between administration dates, it was not possible, except for a few mathematical questions, to attribute these changes to curricular or societal factors.

APPENDIX A:

INSTRUCTIONS TO PANELS AND RATING FORM

INSTRUCTIONS FOR REVIEWING VERBAL TEST QUESTIONS

Educational Testing Service is conducting a study on behalf of the College Board to determine whether or not certain questions that have been administered in two previous editions of the Scholastic Aptitude Test (SAT) represent different tasks when given at different points in time. It is conceivable that, due to curricular or social changes, some questions became more (or less) difficult between the first and the second test administrations.

Consider, for example, how two reading selections, one about the blitzkriegs of the Second World War and one about the problems of weightlessness in space, would have reversed as far as familiarity of content is concerned in the last thirty years. It is also conceivable that some reading selections and questions based on them might be as appropriate today as they were in the past, but changing attitudes or new knowledge may change the attractiveness of incorrect multiple-choice alternatives and thus affect the difficulty of the questions.

With the passing of time, certain words may change in meaning or become obsolete. Other words may become more familiar due to more frequent use. Consider this analogy item:

CISTERN:WATER:: (A) shower:cloud (B) official:power
(C) science:matter (D) museum:antiques (E) vault:valuables

It seems likely that the difficulty of this question would be greater today than in the past because fewer of today's students would know the word "cistern."

For each question in the enclosed test booklets, you are to judge whether or not students would have found the question easier or harder at the second administration than at the first administration. The administration dates are given on the cover of the test booklet. You are asked to consider the entire question, including the stem, the alternatives, and the material on which the question is based. In making each judgment, you should assume that the earlier and later student groups were equal in ability and that the change you note in difficulty level was due not to any change in the ability of the students, but to a change in the meaning of the question or the familiarity of the material on which it is based.

Please use the following codes to indicate your judgment about the change in difficulty for each question:

+ + Denotes questions that you think should be considerably easier (much higher percentage of students respond correctly) at the more recent administration date than at the earlier administration date (dates given on test copy).

+ Denotes questions that you think should be somewhat easier at the more recent administration date than at the earlier administration date.

0 Denotes questions that you think should be approximately the same difficulty at the more recent administration date as at the earlier administration date.

- Denotes questions that you think should be somewhat more difficult at the more recent administration date than at the earlier administration date.

- - Denotes questions that you think should be considerably more difficult (much lower percentage of students respond correctly) at the more recent administration date than at the earlier administration date.

Use the Review Form to record your responses. A set of 4 Review Forms are stapled together to correspond to each test booklet. An extra set is included. Be sure to write the Test Number (given on the cover of the test booklet) in the space provided on the Review Form. Columns are included on the Review Form for you to check the appropriate code. Columns are also provided for you to indicate the extent to which curricular changes may have contributed to changes in difficulty during the time between the two administration dates. Leave these columns blank if you do not know whether curricular changes affected the question you are rating. The "Comments" section allows you to indicate the specific reasons (curricular or societal) why you think the difficulty level of a question changed. It is particularly important that a brief remark be made for questions coded either + + or - - .

Before you begin your review, you should scan the enclosed pamphlet "About the SAT." It will probably be easiest for you to make the judgments asked for if you take the tests yourself and make your judgments as you go. Answer keys are provided for your reference.

After you have completed the Review Form,

- Check to make sure you have written in the Test Number, and
- Sign your name in the upper right-hand corner of each Review Form.

Then

- Fill out the social security form (needed by the ETS Payroll Department);
- Return the social security form, test booklets, answer keys, and rating forms in the envelope provided. You may wish to copy your rating form before returning it.

Thank you for your help and cooperation!

INSTRUCTIONS FOR REVIEWING MATHEMATICAL TEST QUESTIONS

Educational Testing Service is conducting a study on behalf of the College Board to determine whether or not certain questions that have been administered in previous editions of the Scholastic Aptitude Test (SAT) represent different tasks when given at different points in time. It is conceivable that, due to curricular or social changes, some questions became more (or less) difficult between the first and the second test administrations.

For example, those questions that are more in the spirit of contemporary mathematics might be expected to be easier for more recent groups of secondary school students. However, those questions that refer to topics or settings that were more familiar to students in the mid-1960's than to recent groups might have become more difficult.

For each question in the enclosed test booklets, you are to judge whether or not students would have found the question easier or harder at the second administration than at the first administration. The administration dates are given on the cover of the test booklet. You are asked to consider the entire question, including the stem, the alternatives, and the material on which the question is based. In making each judgment, you should assume that the earlier and later student groups were equal in ability and that the change you note in difficulty level was due not to any change in the ability of the students, but to a change in the meaning of the question or the familiarity of the material on which it is based.

Please use the following codes to indicate your judgment about the change in difficulty for each question:

+ + Denotes questions that you think should be considerably easier (much higher percentage of students respond correctly) at the more recent administration date than at the earlier administration date (dates given on test copy).

+ Denotes questions that you think should be somewhat easier at the more recent administration date than at the earlier administration date.

0 Denotes questions that you think should be approximately the same difficulty at the more recent administration date as at the earlier administration date.

- Denotes questions that you think should be somewhat more difficult at the more recent administration date than at the earlier administration date.

- - Denotes questions that you think should be considerably more difficult (much lower percentage of students respond correctly) at the more recent administration date than at the earlier administration date.

Use the Review Form to record your responses. Fill out one Review Form for each test booklet. Be sure to write the Test Number (given on the cover of the test booklet) in the space provided on the Review Form. Columns are included on the Review Form for you to check the appropriate code. Columns are also provided for you to indicate the extent to which curricular changes may have contributed to changes in difficulty during the time between the two administration dates. Leave these columns blank if you do not know whether curricular changes affected the question you are rating. The "Comments" section allows you to indicate the specific reasons (curricular or societal) why you think the difficulty level of a question changed. It is particularly important that a brief remark be made for questions coded either + + or - - .

Before you begin your review, you should scan the enclosed pamphlet "About the SAT." It will probably be easiest for you to make the judgments asked for if you take the tests yourself and make your judgments as you go. Answer keys are provided for your reference.

After you have completed the Review Form,

- Check to make sure you have written in the Test Number, and
- Sign your name in the upper right-hand corner of each Review Form.

Then

- Fill out the social security form (needed by the ETS Payroll Department);
- Return the social security form, test booklets, answer keys, and rating forms in the envelope provided. You may wish to copy your rating form before returning it.

Thank you for your help and cooperation!

REVIEW FORM

Rater ____________________

Test No. ____________________

Question Number	Difficulty for Recent Group (Check One)					Effect of Curricular Changes on the Judged Change in Difficulty Over Time (Check One)			COMMENTS (This space can be used to explain your coded response. It is particularly important that a brief remark be made for questions coded either + + or - - .)
	Easier		=	Harder		Considerable Effect	Some Effect	Little or No Effect	
	++	+	0	-	--				
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									

APPENDIX B:

DATA, TABLES, AND FIGURES

Table B1. Ratings on Verbal Questions

Question	Verbal test 1 rater							Verbal test 2 rater						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	5	3	3	3	3	3	3	3	3	3	3	3	3	2
2	2	2	1	3	2	3	2	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	4	3	3	3	3	3	2
4	3	3	3	2	3	3	3	2	3	1	3	3	3	3
5	1	2	3	3	3	3	1	2	3	2	3	3	3	3
6	2	3	3	3	2	3	1	3	3	3	3	3	3	3
7	1	3	2	3	3	3	2	5	3	2	3	3	3	3
8	2	3	3	3	3	3	1	3	3	2	3	3	3	3
9	5	3	1	3	3	3	3	3	3	2	3	3	3	3
10	4	3	2	4	3	2	1	2	3	1	3	3	3	3
11	3	3	2	3	3	3	1	2	3	3	2	3	3	3
12	3	3	2	3	3	3	3	4	3	1	3	3	3	2
13	2	3	2	3	3	2	3	4	3	2	3	3	3	3
14	2	3	3	2	3	3	3	4	3	3	3	3	3	3
15	1	3	2	3	3	3	3	4	3	1	3	3	3	3
16	1	3	1	2	3	3	1	4	3	3	3	3	3	3
17	4	3	1	3	3	3	3	4	3	3	3	3	3	3
18	1	3	1	3	3	3	1	3	3	2	3	3	3	3
19	2	3	1	3	3	3	1	4	3	2	2	3	2	3
20	1	3	1	3	3	3	1	3	3	1	3	3	3	3
21	3	3	3	3	3	3	3	3	3	3	3	3	3	3
22	3	3	3	3	3	3	2	2	3	2	3	3	3	3
23	2	3	3	3	3	3	3	4	3	2	2	3	3	3
24	3	3	3	3	3	3	3	4	3	2	3	3	3	3
25	2	3	2	2	3	3	3	2	3	1	3	3	3	2
26	4	2	2	3	3	3	3	4	3	3	3	3	3	3
27	2	4	3	2	3	3	4	4	3	2	3	3	3	3
28	4	3	1	3	3	3	2	3	3	2	3	3	3	3

(Continued)

Question	Verbal test 1 rater							Verbal test 2 rater						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
29	3	3	3	2	3	3	1	1	3	1	3	3	3	3
30	5	3	3	3	3	3	1	2	3	2	3	3	3	3
31	2	3	3	3	3	3	2	4	3	1	3	3	3	3
32	3	3	2	2	3	3	2	3	3	1	3	3	3	3
33	3	3	3	3	3	3	2	3	3	2	3	3	3	3
34	1	3	3	3	3	3	2	4	3	2	3	3	3	3
35	2	3	3	3	3	2	2	4	3	2	3	3	3	3
36	3	3	1	3	3	3	3	3	3	1	3	3	3	3
37	2	3	3	3	3	3	3	2	3	1	3	3	3	3
38	2	3	2	3	3	3	3	2	3	1	3	3	3	3
39	3	3	3	3	3	3	3	2	3	1	2	3	3	3
40	3	3	3	3	3	3	3	2	3	1	2	3	3	3
41	4	3	3	3	3	3	3	4	3	3	3	3	3	3
42	5	3	3	3	3	3	3	4	3	3	3	3	3	3
43	4	3	3	3	3	3	3	4	3	3	3	3	3	3
44	4	3	1	3	3	3	3	4	3	3	3	3	3	3
45	2	3	3	3	3	3	3	3	3	3	3	3	3	2
46	2	3	1	2	3	2	2	2	3	2	3	3	3	3
47	4	3	3	3	3	2	3	3	3	3	3	3	3	3
48	4	3	3	3	3	2	3	2	3	1	3	3	3	3
49	4	3	3	3	3	2	2	3	3	3	3	3	3	3
50	5	3	3	3	3	2	3	3	3	2	3	3	3	3
51	3	3	3	3	3	3	3	4	3	3	3	3	3	3
52	4	3	2	3	3	3	3	5	3	3	3	3	4	3
53	5	3	3	3	3	3	3	3	3	3	3	3	3	3
54	1	3	3	3	3	3	2	4	3	2	3	3	3	1
55	5	3	3	3	3	3	4	2	3	2	3	3	3	2
56	2	3	3	3	3	3	3	4	3	2	3	3	3	3
57	2	3	1	3	3	3	1	2	3	2	3	3	3	3
58	3	3	1	1	3	3	1	2	3	2	3	3	3	3
59	4	3	3	3	3	3	3	3	3	3	3	3	3	3
60	2	3	3	3	3	3	3	3	3	3	3	3	3	3
61	2	3	2	3	3	3	3	2	3	3	3	3	3	3
62	2	3	1	3	3	3	3	2	3	2	3	3	3	3
63	1	3	1	2	3	3	3	3	3	2	3	3	3	3

(Continued)

Table B1 (Continued)

	Verbal test 1 rater							Verbal test 2 rater						
Question	1	2	3	4	5	6	7	1	2	3	4	5	6	7
64	2	3	3	3	3	3	2	3	3	2	3	3	3	3
65	4	3	3	3	3	3	3	2	3	1	3	3	3	2
66	1	3	2	3	3	3	2	1	3	2	3	3	3	2
67	3	3	3	3	3	3	2	2	3	3	3	3	3	3
68	3	3	3	3	3	3	3	3	3	2	3	3	3	3
69	4	3	3	3	3	3	3	2	3	2	3	3	2	2
70	2	3	3	3	3	3	3	3	3	1	2	3	3	1
71	2	3	3	3	3	3	2	2	3	2	3	3	3	3
72	4	3	2	1	3	3	1	2	3	3	3	3	3	3
73	4	3	2	1	3	3	1	2	3	1	3	3	3	3
74	3	3	1	2	3	3	1	2	3	1	3	3	3	3
75	3	3	3	3	3	3	3	1	3	1	2	3	3	3
76	2	3	1	2	3	3	1	2	3	1	3	3	3	3
77	2	3	3	2	3	3	1	2	3	1	3	3	3	3
78	2	3	2	2	3	3	1	2	3	2	2	3	3	3
79	2	3	2	2	3	3	1	2	3	2	3	3	3	3
80	2	3	2	1	3	3	1	2	3	1	2	3	3	3
81	4	3	3	3	3	3	2	4	3	2	3	3	3	3
82	3	3	2	3	3	3	2	2	3	2	3	3	3	3
83	4	3	2	3	3	3	2	4	3	3	3	3	3	3
84	4	3	3	2	3	3	2	4	3	2	3	3	3	3
85	4	3	1	2	3	3	2	4	3	2	3	3	3	1
86	4	4	2	3	3	4	4	2	3	2	3	3	3	3
87	5	3	2	5	3	4	4	2	3	2	3	3	3	3
88	2	3	2	4	3	4	4	2	3	1	3	3	3	3
89	2	3	2	5	3	4	4	2	3	2	3	3	3	3
90	4	3	2	4	3	4	4	3	3	2	3	3	3	2
Mean	2.87	2.99	2.32	2.81	2.98	2.97	2.38	2.89	3.00	2.02	2.89	3.00	2.99	2.82
S.D.	1.17	0.24	0.79	0.66	0.15	0.38	0.93	0.94	0.00	0.75	0.31	0.00	0.18	0.46

Table B2. Ratings on Mathematical Questions

	Mathematics test 1 rater							Mathematics test 2 rater						
Question	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	2	3	3	3	3	3	3	4	3	3	3	3	3	3
2	4	3	3	3	3	3	3	2	2	3	2	2	3	2
3	2	3	4	3	3	3	3	4	4	3	3	3	3	3
4	3	4	3	3	3	3	4	4	4	3	2	3	3	3
5	2	3	3	3	4	3	4	4	3	4	5	4	3	3
6	2	3	3	3	3	3	3	3	4	3	2	3	3	3
7	4	4	3	4	4	3	3	2	2	4	3	3	3	3
8	4	2	3	3	3	3	3	2	3	4	3	3	3	3
9	2	3	3	3	3	3	3	4	3	3	4	4	3	4
10	2	2	3	2	3	3	2	4	5	4	2	3	3	3
11	3	2	4	3	3	3	3	5	5	4	5	4	4	4
12	2	2	3	2	3	3	3	3	4	3	3	3	3	3
13	3	5	3	3	3	3	3	4	3	2	2	3	3	3
14	5	3	3	3	3	3	2	4	4	4	4	3	3	4
15	2	3	3	3	3	3	3	3	3	4	3	3	3	3
16	3	4	3	3	3	3	3	4	4	3	4	3	4	3
17	2	1	3	2	2	3	3	4	5	4	5	3	4	4
18	3	4	3	3	3	3	3	4	5	4	5	3	4	4
19	2	5	4	3	2	3	3	4	4	4	4	3	3	3
20	1	2	4	3	3	3	2	2	4	4	4	3	3	3
21	4	2	3	3	2	3	3	4	4	3	4	3	3	4
22	3	1	3	3	3	3	3	3	4	3	3	3	3	3
23	2	2	4	2	3	3	2	3	4	3	5	3	4	3
24	4	4	3	3	2	3	3	4	3	3	3	3	2	4
25	3	3	3	3	3	3	3	3	3	4	4	3	3	4
Mean	2.76	2.92	3.20	2.88	2.92	3.00	2.92	3.48	3.68	3.44	3.48	3.08	3.16	3.28
S.D.	0.95	1.06	0.40	0.43	0.48	0.00	0.48	0.81	0.84	0.57	1.02	0.39	0.46	0.53

Table B2 (Continued)

Question	Mathematics test 3 rater 1	2	3	4	5	6	7	Mathematics test 4 rater 1	2	3	4	5	6	7
1	3	3	3	3	3	3	3	3	3	3	4	3	3	3
2	3	3	3	3	3	3	3	3	3	4	3	3	3	3
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	3	4	3	3	3	3	2	3	3	3	3	3	3
5	3	3	4	3	3	3	3	3	3	4	3	3	3	3
6	3	2	3	3	3	3	3	4	3	3	3	3	3	3
7	2	4	4	3	3	3	4	3	3	4	4	3	3	3
8	4	4	4	3	2	3	3	3	3	3	3	3	3	3
9	3	3	4	3	3	3	3	3	3	4	3	4	3	3
10	3	3	4	3	3	3	3	3	3	4	3	3	3	3
11	4	3	3	3	3	3	4	3	3	3	3	4	3	3
12	3	3	4	2	4	4	4	3	3	3	3	4	3	3
13	3	3	3	3	3	3	3	3	3	4	2	3	4	3
14	3	3	4	2	3	3	3	2	3	4	3	2	3	3
15	3	3	3	3	3	3	3	3	3	3	3	3	3	3
16	3	3	4	2	3	3	4	3	3	4	3	3	3	3
17	3	3	3	3	3	3	3	3	3	3	3	3	3	3
18	2	3	3	3	3	4	4	3	3	4	3	3	3	3
19	3	3	3	3	3	3	3	3	3	2	3	3	3	3
20	3	3	3	3	3	3	3	3	3	4	3	3	3	3
21	3	2	3	3	3	3	4	3	3	4	3	3	3	3
22	3	3	4	3	3	3	4	2	3	3	3	3	3	2
23	2	3	4	3	2	3	3	2	3	4	3	3	3	3
24	3	2	3	3	2	3	3	2	3	3	3	3	3	3
25	1	2	3	3	3	3	3	3	3	3	3	2	3	3
Mean	2.92	2.92	3.44	2.88	2.92	3.08	3.28	2.84	3.00	3.44	3.04	3.04	3.04	2.96
S.D.	0.63	0.48	0.50	0.33	0.39	0.27	0.45	0.46	0.00	0.57	0.34	0.45	0.20	0.20

238

239

Table B3. Rater Means and Standard Deviations

	Verbal					
	Test 1		Test 2		Total	
Rater	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	2.87	1.17	2.89	0.94	2.88	1.06
2	2.99	0.24	3.00	0.00	2.99	0.17
3	2.32	0.79	2.02	0.75	2.17	0.78
4	2.81	0.66	2.89	0.31	2.85	0.52
5	2.98	0.15	3.00	0.00	2.99	0.10
6	2.97	0.38	2.99	0.18	2.98	0.30
7	2.38	0.93	2.82	0.46	2.60	0.76
Ave.	2.76	0.35	2.80	0.23	2.78	0.30

	Mathematics					
	Test 1		Test 2		Test 3	
Rater	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	2.76	0.95	3.48	0.81	2.92	0.63
2	2.92	1.06	3.68	0.84	2.92	0.48
3	3.20	0.40	3.44	0.57	3.44	0.50
4	2.88	0.43	3.48	1.02	2.88	0.33
5	2.92	0.48	3.08	0.39	2.92	0.39
6	3.00	0.00	3.16	0.46	3.08	0.27
7	2.92	0.48	3.28	0.53	3.28	0.45
Ave.	2.94	0.29	3.37	0.45	3.06	0.19

	Mathematics			
	Test 4		Total	
Rater	Mean	S.D.	Mean	S.D.
1	2.84	0.46	3.00	0.79
2	3.00	0.00	3.13	0.78
3	3.44	0.57	3.38	0.52
4	3.04	0.34	3.07	0.65
5	3.04	0.45	2.99	0.44
6	3.04	0.20	3.07	0.29
7	2.96	0.20	3.11	0.47
Ave.	3.05	0.15	3.11	0.33

Table B4. Intercorrelations

Verbal test 1

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.0436	0.0549	0.1147	0.1362	-0.0373	0.2413	0.6207	0.0983
2	0.0436	1.0000	0.1409	-0.1193	0.3414	-0.0045	0.1940	0.2216	-0.1600
3	0.0549	0.1409	1.0000	0.1915	0.0899	-0.1854	0.1782	0.4568	0.1993
4	0.1147	-0.1193	0.1915	1.0000	-0.0513	0.3192	0.5965	0.6633	0.0808
5	0.1362	0.3414	0.0899	-0.0513	1.0000	-0.0065	0.1825	0.2457	0.1345
6	-0.0373	-0.0045	-0.1854	0.3192	-0.0065	1.0000	0.2388	0.2524	-0.0910
7	0.2413	0.1940	0.1782	0.5965	0.1825	0.2388	1.0000	0.7848	0.0570
Ave.	0.6207	0.2216	0.4568	0.6633	0.2457	0.2524	0.7848	1.0000	0.1346
Distance . . .	0.0983	-0.1600	0.1993	0.0808	0.1545	-0.0910	0.0570	0.1346	1.0000

Verbal test 2

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.0000	0.3270	0.0727	0.0000	0.1565	-0.1425	0.7309	-0.0379
2	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
3	0.3270	0.0000	1.0000	0.2543	0.0000	0.1061	0.0671	0.7520	0.1063
4	0.0727	0.0000	0.2543	1.0000	0.0000	0.2078	0.0829	0.4099	-0.0818
5	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
6	0.1565	0.0000	0.1061	0.2078	0.0000	1.0000	0.1206	0.3519	-0.0656
7	-0.1425	0.0000	0.0671	0.0829	0.0000	0.1206	1.0000	0.2880	0.0228
Ave.	0.7309	0.0000	0.7520	0.4099	0.0000	0.3519	0.2880	1.0000	0.0111
Distance . . .	-0.0379	0.0000	0.1063	-0.0818	0.0000	-0.0656	0.0228	0.0111	1.0000

Table B4 (Continued)

Total verbal

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.0396	0.1562	0.1070	0.1139	0.0228	0.1395	0.6560	0.0623
2	0.0396	1.0000	0.0852	-0.1002	0.3474	-0.0040	0.1861	0.1937	-0.1378
3	0.1562	0.0852	1.0000	0.1716	0.0428	-0.0778	0.0860	0.5282	0.1485
4	0.1070	-0.1002	0.1716	1.0000	-0.0341	0.2933	0.5113	0.6057	0.0586
5	0.1139	0.3474	0.0428	-0.0341	1.0000	-0.0056	0.1841	0.2166	0.1260
6	0.0228	-0.0040	-0.0778	0.2933	-0.0056	1.0000	0.2047	0.2773	-0.0851
7	0.1395	0.1861	0.0860	0.5113	0.1841	0.2047	1.0000	0.6598	0.0624
Ave.	0.6560	0.1937	0.5282	0.6057	0.2166	0.2773	0.6598	1.0000	0.1094
Distance . . .	0.0623	-0.1378	0.1485	0.0586	0.1260	-0.0851	0.0624	0.1094	1.0000

Math test 1

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.2023	-0.1048	0.4450	0.1048	0.0	-0.2774	0.5973	-0.5697
2	0.2023	1.0000	-0.2757	0.5319	0.1654	0.0	0.2917	0.7297	-0.5622
3	-0.1048	-0.2757	1.0000	0.0551	-0.1429	0.0	0.0	-0.0125	0.0510
4	0.4450	0.5319	0.0551	1.0000	0.4962	0.0	0.2917	0.8554	-0.2727
5	0.1048	0.1654	-0.1429	0.4962	1.0000	0.0	0.3780	0.4884	-0.1615
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	-0.2774	0.2917	0.0	0.2917	0.3780	0.0	1.0000	0.3976	0.0370
Ave.	0.5973	0.7297	-0.0125	0.8554	0.4884	0.0	0.3976	1.0000	-0.5991
Distance . . .	-0.5697	-0.5622	0.0510	-0.2727	-0.1615	0.0	0.0370	-0.5991	1.0000

Table B4 (Continued)

Math test 2

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.5651	-0.0930	0.3269	0.5200	0.3231	0.5769	0.6853	-0.6091
2	0.5651	1.0000	0.2469	0.3916	0.1955	0.5488	0.5476	0.7521	-0.0131
3	-0.0930	0.2469	1.0000	0.4352	0.2045	0.1355	0.2613	0.4406	-0.0711
4	0.3269	0.3916	0.4352	1.0000	0.5067	0.6549	0.6133	0.8268	-0.3899
5	0.5200	0.1955	0.2045	0.5067	1.0000	0.1420	0.5380	0.6096	-0.6608
6	0.3231	0.5488	0.1355	0.6549	0.1420	1.0000	0.3915	0.6650	-0.0819
7	0.5769	0.5476	0.2613	0.6133	0.5380	0.3915	1.0000	0.8054	-0.3310
Ave.	0.6853	0.7521	0.4406	0.8268	0.6096	0.6655	0.8054	1.0000	-0.4411
Distance . . .	-0.6091	-0.0131	-0.0711	-0.3899	-0.6608	-0.0819	-0.3310	-0.4411	1.0000

Math test 3

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	-0.0354	-0.0147	0.1157	-0.3644	-0.0625	-0.1387	0.2822	-0.0219
2	-0.0354	1.0000	0.4167	0.0655	-0.4123	-0.0354	0.2550	0.5666	-0.0267
3	-0.0147	0.4167	1.0000	-0.4364	0.0000	0.2357	0.2451	0.6318	-0.1707
4	0.1157	0.0655	-0.4364	1.0000	-0.4499	-0.5401	-0.4708	-0.2264	-0.1185
5	-0.3644	-0.4123	0.0000	-0.4499	1.0000	0.7289	0.4043	0.1645	0.1751
6	-0.0625	-0.0354	0.2357	-0.5401	0.7289	1.0000	0.4507	0.5220	-0.0044
7	-0.1387	0.2550	0.2451	-0.4708	0.4043	0.4507	1.0000	0.6261	0.0552
Ave	0.2822	0.5666	0.6318	-0.2264	0.1645	0.5220	0.6261	1.0000	-0.0666
Distance . . .	-0.0219	-0.0267	-0.1707	-0.1185	0.1751	-0.0044	0.0552	-0.0666	1.0000

Table B4 (Continued)

Math test 4

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.0000	-0.2075	0.0299	0.3158	0.0526	0.0000	0.6227	0.1501
2	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	-0.0000
3	-0.2075	0.0000	1.0000	-0.1439	-0.2536	0.2075	0.0000	0.3616	0.1540
4	0.0299	0.0000	-0.1439	1.0000	-0.0299	-0.6272	0.0000	0.2219	-0.2577
5	0.3158	0.0000	-0.2536	-0.0299	1.0000	-0.0526	0.0000	0.5358	0.4803
6	0.0526	0.0000	0.2075	-0.6272	-0.0526	1.0000	0.0000	0.1014	0.1783
7	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	-0.0000
Ave.	0.6227	0.0000	0.3616	0.2219	0.5358	0.1014	0.0000	1.0000	0.3677
Distance . . .	0.1501	-0.0000	0.1540	-0.2577	0.4803	0.1783	-0.0000	0.3677	1.0000

Total math

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.4284	-0.0336	0.3501	0.2010	0.2651	0.2125	0.6580	-0.4134
2	0.4284	1.0000	0.1038	0.4390	0.0706	0.4000	0.4216	0.7402	-0.1827
3	-0.0336	0.1038	1.0000	0.1625	-0.0366	0.1772	0.1878	0.3441	-0.0158
4	0.3501	0.4390	0.1625	1.0000	0.2524	0.3967	0.3444	0.7461	-0.3168
5	0.2010	0.0706	-0.0366	0.2524	1.0000	0.1684	0.3380	0.4065	-0.1669
6	0.2651	0.4000	0.1772	0.3967	0.1684	1.0000	0.3362	0.5834	-0.0344
7	0.2125	0.4216	0.1878	0.3444	0.3380	0.3362	1.0000	0.6274	-0.1419
Ave.	0.6580	0.7402	0.3441	0.7461	0.4065	0.5834	0.6274	1.0000	-0.3481
Distance . . .	-0.4134	-0.1827	-0.0158	-0.3168	-0.1669	-0.0344	-0.1419	-0.3481	1.0000

Table B5. Verbal Test 1, Question Data

Question	Delta		Distance from major axis	Rater	
	May 1976	May 1963		Mean	S.D.
1	9.0	8.1	0.22	3.29	0.70
2*	11.7	12.1	-0.70	2.14	0.64
3	13.0	11.5	0.64	3.00	0.00
4	11.7	11.4	-0.20	2.86	0.35
5*	12.8	11.8	0.29	2.29	0.88
6*	14.1	13.8	-0.20	2.43	0.73
7*	15.5	14.9	0.01	2.43	0.73
8*	15.6	15.3	-0.20	2.57	0.73
9	15.6	14.9	0.08	3.00	1.07
10**	15.8	16.0	-0.56	2.71	1.03
11*	9.2	8.3	0.22	2.57	0.73
12	10.7	10.9	-0.56	2.86	0.35
13*	12.1	11.6	-0.06	2.57	0.49
14	12.6	11.9	0.08	2.71	0.45
15*	13.7	12.2	0.64	2.57	0.73
16*	15.6	15.6	-0.42	2.00	0.93
17**	17.5	17.2	-0.20	2.86	0.83
18**	16.5	16.4	-0.34	2.14	0.99
19**	16.7	17.5	-0.98	2.29	0.88
20**	17.9	17.9	-0.41	2.14	0.99
21	7.6	8.5	-1.05	3.00	0.00
22**	5.5	6.0	-0.77	2.86	0.35
23	7.4	7.9	-0.77	2.86	0.35
24	10.0	10.6	-0.84	3.00	0.00
25*	11.2	9.8	0.57	2.57	0.49
26	12.5	11.1	0.57	2.86	0.64
27	11.6	12.6	-1.12	3.00	0.76
28	12.7	13.0	-0.63	2.71	0.88
29*	14.5	14.9	-0.70	2.57	0.73
30**	16.2	16.7	-0.77	3.00	1.07
31	11.0	10.1	0.22	2.71	0.45
32*	13.0	12.3	0.08	2.57	0.49
33	10.9	10.7	-0.28	2.86	0.35
34**	16.8	15.9	0.22	2.57	0.73
35*	12.5	12.0	-0.06	2.57	0.49
36	14.9	14.8	-0.34	2.71	0.70
37	15.4	14.3	0.36	2.86	0.35
38**	16.3	15.8	-0.06	2.71	0.45
39	11.5	11.5	-0.42	3.00	0.00
40	13.2	13.3	-0.49	3.00	0.00
41	11.0	8.1	1.63	3.14	0.35
42	11.1	9.0	1.07	3.29	0.70
43	11.7	10.2	0.64	3.14	0.35
44	11.2	10.9	-0.20	2.86	0.83

(Continued)

Table B5 (Continued)

Question	Delta		Distance from major axis	Rater	
	May 1976	May 1963		Mean	S.D.
45	14.1	13.4	0.08	2.86	0.35
46**	16.7	16.5	-0.27	2.14	0.64
47	11.4	10.4	0.29	3.00	0.53
48**	16.5	14.8	0.79	3.00	0.53
49	14.4	14.3	-0.34	2.86	0.64
50	11.5	10.7	0.15	3.14	0.83
51	9.9	9.2	0.08	3.00	0.00
52	10.9	9.9	0.29	3.00	0.53
53	12.3	11.1	0.43	3.29	0.70
54*	12.3	11.0	0.50	2.57	0.73
55*	12.9	12.6	-0.20	3.43	0.73
56	13.3	13.2	-0.35	2.86	0.35
57**	16.2	15.6	0.01	2.29	0.88
58*	15.5	15.7	-0.56	2.14	0.99
59	7.9	6.8	0.36	3.14	0.35
60	12.5	9.9	1.42	2.86	0.35
61	12.9	12.2	0.08	2.71	0.45
62*	12.9	13.1	-0.56	2.57	0.73
63*	14.6	15.0	-0.70	2.29	0.88
64	15.5	15.2	-0.20	2.71	0.45
65**	17.5	16.8	0.08	3.14	0.35
66**	16.4	17.5	-1.19	2.43	0.73
67	7.6	7.4	-0.28	2.86	0.35
68	8.3	8.1	-0.28	3.00	0.00
69	9.0	8.7	-0.21	3.14	0.35
70	11.4	10.9	-0.06	2.86	0.35
71	11.4	11.1	-0.20	2.71	0.45
72*	12.9	12.8	-0.35	2.43	1.05
73*	13.8	13.3	-0.06	2.43	1.05
74*	15.3	15.0	-0.20	2.29	0.88
75	15.0	13.1	0.93	3.00	0.00
76*	12.9	11.7	0.43	2.14	0.83
77*	13.1	12.5	0.01	2.43	0.73
78*	14.4	14.0	-0.13	2.29	0.70
79**	17.0	16.7	-0.20	2.29	0.70
80*	14.5	13.8	0.08	2.14	0.83
81	12.3	12.2	-0.35	3.00	0.53
82	13.7	13.4	-0.20	2.71	0.45
83	12.4	11.6	0.15	2.86	0.64
84	13.6	13.3	-0.20	2.86	0.64
85*	15.5	15.0	-0.06	2.57	0.90
86**	16.5	15.8	0.08	3.43	0.73
87*	11.3	11.4	-0.49	3.71	1.03
88	13.9	13.8	-0.35	3.14	0.83

(Continued)

Table B5 (Continued)

Question	Delta May 1976	 May 1963	Distance from major axis	Rater Mean	 S.D.
89	14.7	13.9	0.15	3.29	1.03
90*	11.0	11.1	-0.49	3.43	0.73
Delta mean	13.1	12.6	Rater mean	2.76	
Delta S.D.	2.6	2.7	Rater S.D.	0.35	

*Questions excluded from delta plots because rater mean is less than 2.7 or greater than 3.3.

**Questions excluded from delta plots and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table B6. Verbal Test 2, Question Data

Question	Delta May 1976	 Mar. 1973	Distance from major axis	Rater Mean	 S.D.
1	9.8	9.6	0.46	2.86	0.35
2	10.9	11.0	0.20	3.00	0.00
3	13.3	13.6	-0.03	3.00	0.53
4*	14.1	14.6	-0.21	2.57	0.73
5**	16.0	16.5	-0.28	2.71	0.45
6	8.1	9.0	-0.27	3.00	0.00
7	11.7	12.5	-0.33	3.14	0.83
8	14.8	14.9	0.06	2.86	0.35
9**	15.8	16.1	-0.12	2.86	0.35
10**.	16.7	16.2	0.42	2.57	0.73
11	8.7	9.9	-0.51	2.71	0.45
12	10.2	12.0	-1.00	2.71	0.88
13	12.6	13.1	-0.15	3.00	0.53
14	15.5	15.6	0.03	3.14	0.35
15**.	15.8	16.0	-0.05	2.86	0.83
16	10.0	10.4	0.02	3.14	0.35
17	11.5	12.0	-0.11	3.14	0.35
18	11.9	12.4	-0.12	2.86	0.35
19	14.4	14.4	0.15	2.71	0.70
20**.	16.0	16.2	-0.06	2.71	0.35
21	9.3	9.8	-0.03	3.00	0.00
22	14.3	14.6	-0.07	2.71	0.45

(Continued)

Table B6 (Continued)

Question	Delta		Distance from major axis	Rater	
	May 1976	Mar. 1973		Mean	S.D.
23	15.7	15.0	0.60	2.86	0.64
24	14.0	14.4	-0.13	3.00	0.53
25**	17.5	17.5	0.03	2.43	0.73
26	9.2	9.9	-0.17	3.14	0.35
27	11.6	12.4	-0.33	3.00	0.53
28	13.8	13.6	0.31	2.86	0.35
29*	15.2	15.7	-0.25	2.43	0.90
30**	16.8	17.7	-0.60	2.71	0.45
31	10.4	10.2	0.44	2.86	0.83
32	14.3	14.4	0.08	2.71	0.70
33	10.2	10.8	-0.13	2.86	0.35
34	9.9	10.5	-0.12	3.00	0.53
35	15.2	15.3	0.04	3.00	0.53
36	12.8	12.9	0.13	2.71	0.70
37*	14.2	14.2	0.15	2.57	0.73
38*	14.8	15.0	-0.01	2.57	0.73
39**	16.0	16.2	-0.06	2.43	0.73
40*	14.1	14.7	-0.28	2.43	0.73
41	9.0	9.2	0.20	3.14	0.35
42	11.8	12.3	-0.12	3.14	0.35
43	12.5	12.6	0.14	3.14	0.35
44	13.8	13.7	0.24	3.14	0.35
45	11.1	11.4	0.05	2.86	0.35
46	11.9	12.3	-0.05	2.71	0.45
47	13.1	13.2	0.12	3.00	0.00
48*	13.3	13.4	0.11	2.57	0.73
49	10.8	11.1	0.06	3.00	0.00
50	12.0	12.6	-0.20	2.86	0.35
51**	6.0	7.4	-0.56	3.14	0.35
52*	8.3	9.0	-0.14	3.43	0.73
53	9.3	9.0	0.55	3.00	0.00
54	14.8	15.2	-0.16	2.71	0.88
55**	15.7	16.0	-0.12	2.57	0.49
56	14.8	14.8	0.13	3.00	0.53
57**	17.7	17.1	0.46	2.71	0.45
58**	16.5	17.3	-0.51	2.71	0.45
59	6.8	6.8	0.43	3.00	0.00
60	9.9	10.0	0.24	3.00	0.00
61	11.5	11.8	0.04	2.86	0.35
62	12.6	12.7	0.14	2.71	0.45
63	13.1	13.1	0.19	2.86	0.35
64	13.8	14.5	-0.34	2.86	0.35
65**	16.5	16.0	0.43	2.43	0.73
66**	17.1	17.7	-0.39	2.43	0.73

(Continued)

Table B6 (Continued)

Question	Delta May 1976	Delta Mar. 1973	Distance from major axis	Rater Mean	Rater S.D.
67	8.5	8.9	0.07	2.86	0.35
68	9.2	10.2	-0.39	2.86	0.35
69*	9.5	10.0	-0.04	2.43	0.49
70*	10.5	10.3	0.44	2.29	0.88
71	12.5	12.9	-0.07	2.71	0.45
72	12.3	13.0	-0.28	2.86	0.35
73**	16.4	16.1	0.29	2.57	0.73
74**	16.8	17.0	-0.09	2.57	0.73
75**	17.7	18.2	-0.34	2.29	0.88
76*	11.5	11.6	0.18	2.57	0.73
77*	11.0	11.3	0.05	2.57	0.73
78**	16.4	16.6	-0.07	2.57	0.49
79	14.2	14.4	0.01	2.71	0.45
80*	13.0	13.4	-0.09	2.43	0.73
81	13.9	14.1	0.02	3.00	0.53
82	13.9	14.2	-0.05	2.71	0.45
83	12.2	12.3	0.15	3.14	0.35
84	14.3	15.1	-0.43	3.00	0.53
85	12.2	12.6	-0.06	2.71	0.88
86	13.5	13.5	0.18	2.71	0.45
87	13.7	13.7	0.17	2.71	0.45
88*	11.6	11.7	0.18	2.57	0.73
89	14.7	14.8	0.06	2.71	0.45
90	13.7	14.0	-0.05	2.71	0.45
Delta mean	12.9	13.2	Rater mean	2.80	
Delta S.D.	2.7	2.5	Rater S.D.	0.23	

*Questions excluded from delta plots because rater mean is less than 2.7 or greater than 3.3.

**Questions excluded from delta plots and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table B7. Mathematics Test 1, Question Data

Question	Delta Jan. 1971	Delta May 1962	Distance from major axis	Rater Mean	Rater S.D.
1	8.9	8.3	-0.09	2.86	0.35
2**	7.4	6.0	0.50	3.14	0.35
3	11.1	10.3	0.02	3.00	0.53
4	9.9	9.9	-0.53	3.29	0.45
5	9.0	8.1	0.12	3.14	0.64
6	13.1	11.9	0.28	2.86	0.35
7*	11.9	11.8	-0.49	3.57	0.49
8	13.2	11.9	0.35	3.00	0.53
9	14.0	12.1	0.76	2.86	0.35
10*	12.9	11.9	0.13	2.43	0.49
11	13.5	12.6	0.05	3.00	0.53
12*	15.5	14.5	0.10	2.57	0.49
13	14.5	14.7	-0.75	3.29	0.70
14	13.8	13.9	-0.66	3.14	0.83
15	14.8	13.4	0.39	2.86	0.35
16	14.5	13.6	0.04	3.14	0.35
17**	16.5	16.7	-0.78	2.29	0.70
18	15.4	14.5	0.03	3.14	0.35
19**	16.0	15.5	-0.27	3.14	0.99
20**	16.8	15.8	0.08	2.57	0.90
21**	16.2	16.9	-1.13	2.86	0.64
22**	16.4	16.7	-0.85	2.71	0.70
23**	16.1	16.5	-0.91	2.57	0.73
24**	15.4	16.0	-1.05	3.14	0.64
25**	17.4	17.5	-0.72	3.00	0.00
Delta mean	13.8	13.2	Rater mean	2.94	
Delta S.D.	2.7	3.0	Rater S.D.	0.29	

*Questions excluded from delta plots because rater mean is less than 2.7 or greater than 3.3.

**Questions excluded from delta plots and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table BS. Mathematics Test 2, Question Data

Question	Delta		Distance from major axis	Rater	
	Dec. 1973	Jan. 1965		Mean	S.D.
1	6.8	7.1	-0.92	3.14	0.35
2*	10.4	8.5	0.89	2.29	0.45
3	9.5	9.0	-0.11	3.29	0.45
4	10.5	9.8	0.12	3.14	0.64
5*	8.5	9.3	-1.07	3.71	0.70
6	10.5	9.2	0.51	3.00	0.53
7	11.2	10.9	-0.06	2.86	0.64
8	9.8	7.9	0.83	3.00	0.53
9*	11.0	11.0	-0.28	3.57	0.49
10*	12.4	11.8	0.26	3.43	0.90
11*	10.6	11.5	-0.91	4.43	0.49
12	13.0	12.6	0.19	3.14	0.35
13	13.2	13.8	-0.44	2.86	0.64
14*	13.2	13.9	-0.50	3.71	0.35
15	14.5	15.3	-0.43	3.14	0.35
16*	14.2	14.1	0.12	3.57	0.49
17*	13.2	13.1	0.02	4.14	0.64
18*	15.4	15.4	0.19	4.14	0.64
19*	13.6	13.7	-0.07	3.57	0.49
20	14.7	14.4	0.31	3.29	0.70
21*	15.1	15.0	0.22	3.57	0.49
22**	16.1	15.1	0.91	3.14	0.35
23*	14.8	15.1	-0.07	3.57	0.73
24**	15.9	15.5	0.50	3.14	0.64
25**	16.8	17.1	0.14	3.43	0.49
Delta mean	12.6	12.4	Rater mean	3.37	
Delta S.D.	2.6	2.8	Rater S.D.	0.45	

*Questions excluded from delta plot because rater mean is less than 2.7 or greater than 3.3.

**Questions excluded from delta plot and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table B9. Mathematics Test 3, Question Data

Question	Delta Nov. 1975	Delta Apr. 1971	Distance from major axis	Rater Mean	Rater S.D.
1	9.3	7.9	0.69	3.00	0.00
2	8.7	8.1	0.12	3.00	0.00
3	9.5	9.5	-0.27	3.00	0.00
4	13.0	12.3	0.31	3.29	0.45
5	9.8	10.3	-0.61	3.14	0.35
6	10.8	10.7	-0.16	2.86	0.35
7	11.5	11.1	0.06	3.29	0.70
8	9.2	9.2	-0.28	3.29	0.70
9	12.3	12.0	0.02	3.14	0.35
10	13.0	12.9	-0.10	3.14	0.35
11	13.5	13.3	-0.02	3.29	0.45
12*	14.0	13.8	-0.01	3.43	0.73
13	13.8	13.6	-0.01	3.00	0.00
14	11.9	11.4	0.14	3.00	0.53
15	14.7	14.3	0.15	3.00	0.00
16	11.6	11.2	0.07	3.14	0.64
17**	15.9	15.8	-0.02	3.00	0.00
18**	16.2	15.7	0.26	3.14	0.64
19	14.8	14.8	-0.12	3.00	0.00
20**	15.4	16.0	-0.52	3.00	0.00
21**	15.7	16.1	-0.37	3.00	0.53
22**	16.1	16.2	-0.15	3.29	0.45
23**	16.2	16.3	-0.15	2.86	0.64
24**	16.8	15.8	0.63	2.71	0.45
Delta mean	13.1	12.8	Rater mean	3.06	
Delta S.D.	2.5	2.6	Rater S.D.	0.19	

*Questions excluded from delta plot because rater mean is less than 2.7 or greater than 3.3.

**Questions excluded from delta plot and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table B10. Mathematics Test 4, Question Data

Question	Delta		Distance from major axis	Rater	
	Apr. 1976	Mar. 1972		Mean	S.D.
1	8.1	8.3	-0.10	3.14	0.35
2	8.7	8.3	0.32	3.14	0.35
3	8.7	9.0	-0.17	3.00	0.00
4	10.2	10.0	0.17	2.86	0.35
5	10.5	10.7	-0.11	3.14	0.35
6	10.0	10.2	-0.11	3.14	0.35
7	11.6	12.0	-0.26	3.29	0.45
8	10.3	10.8	-0.32	3.00	0.00
9	13.5	13.1	0.30	3.29	0.45
10	12.8	12.2	0.45	3.14	0.35
11	14.9	14.7	0.16	3.14	0.35
12	13.3	13.1	0.16	3.14	0.35
13	14.1	13.8	0.23	3.14	0.64
14	14.5	15.4	-0.62	2.86	0.64
15	12.6	13.9	-0.19	3.00	0.00
16	14.7	13.9	0.58	3.14	0.35
17	14.9	15.0	-0.06	3.00	0.00
18	14.7	15.4	-0.48	3.14	0.35
19**	14.8	16.0	-0.83	2.86	0.35
20	15.4	15.3	0.08	3.14	0.35
21**	15.8	16.1	-0.20	3.14	0.35
22**	18.2	18.3	-0.07	2.71	0.45
23	13.9	14.0	-0.05	3.00	0.53
24**	16.9	16.2	0.51	2.86	0.35
Delta mean	13.0	13.1	Rater mean	3.05	
Delta S.D.	2.7	2.7	Rater S.D.	0.15	

*Questions excluded from delta plot because rater mean is less than 2.7 or greater than 3.3.

**Questions excluded from delta plot and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Figure B1. Relationship between Delta and Percent Correct

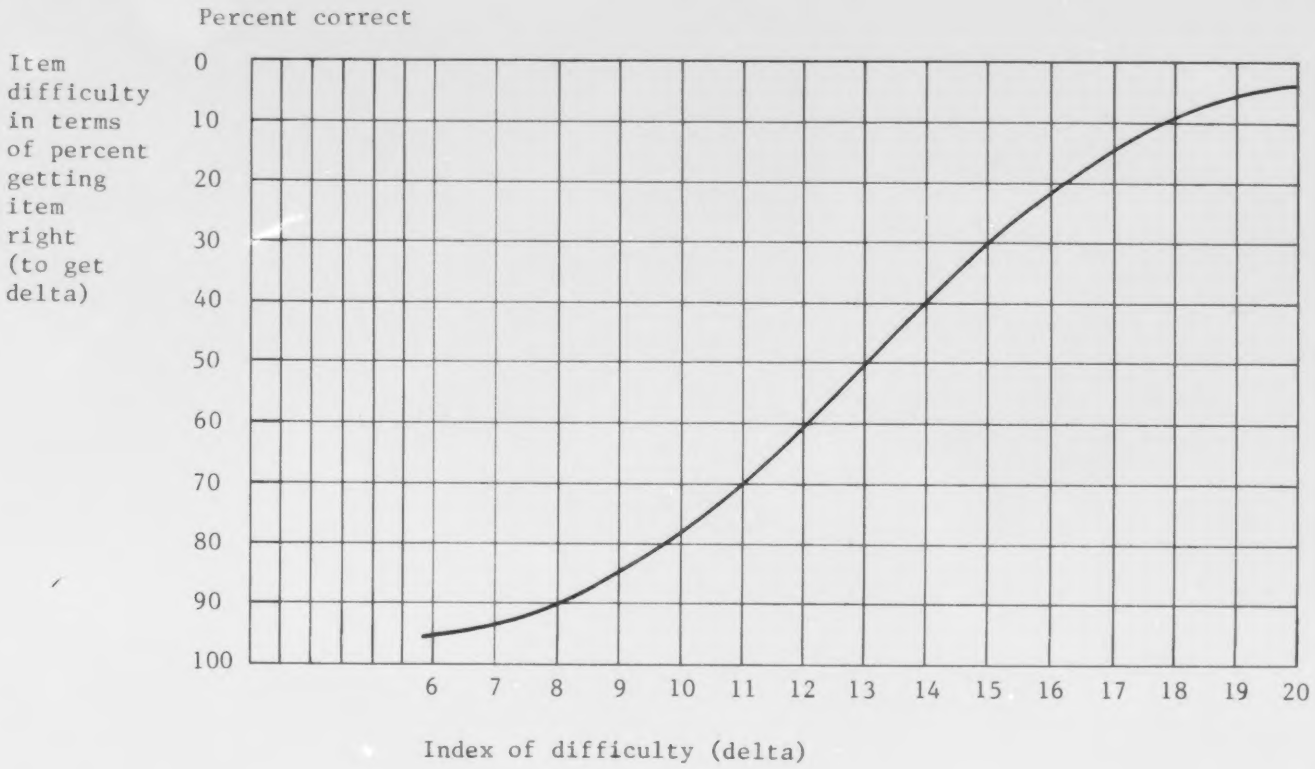
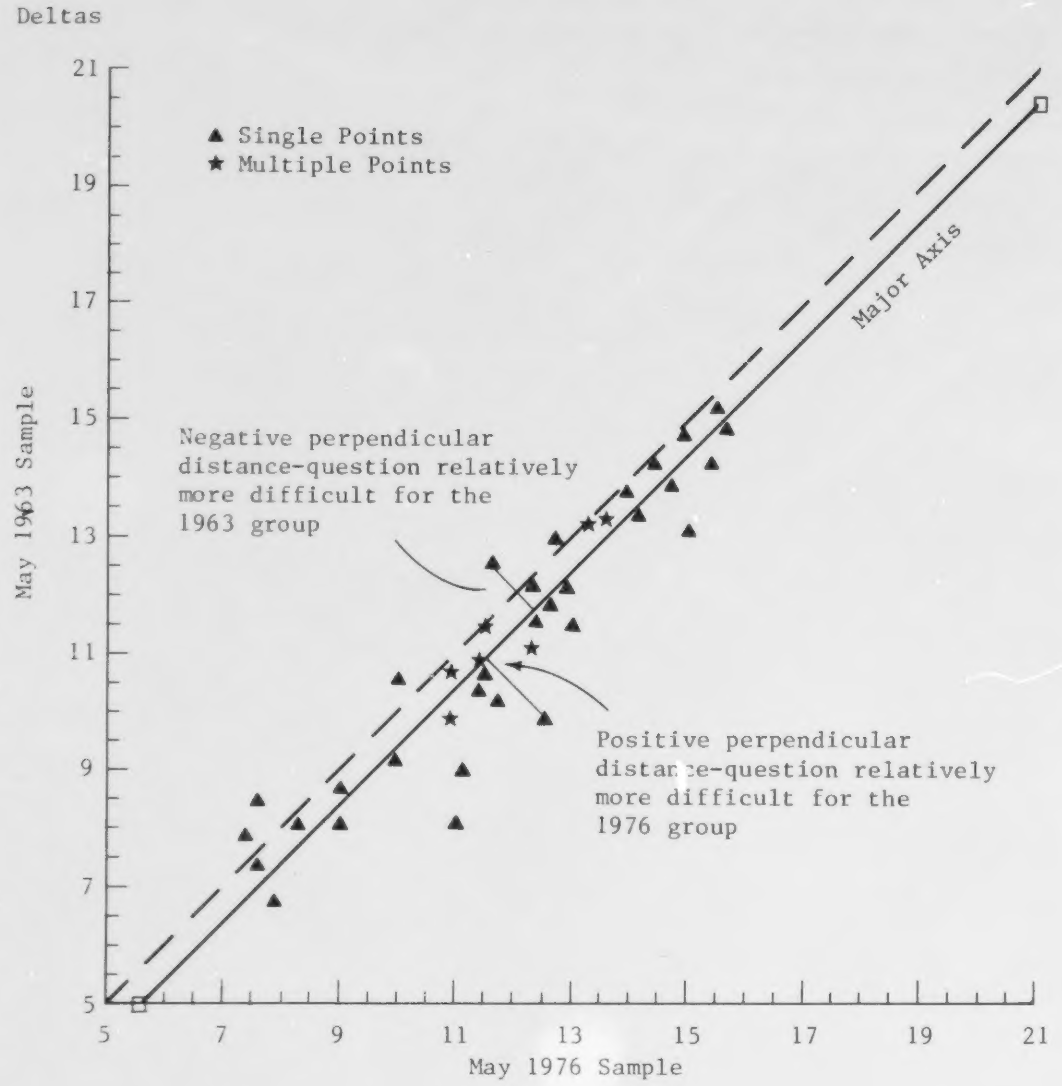
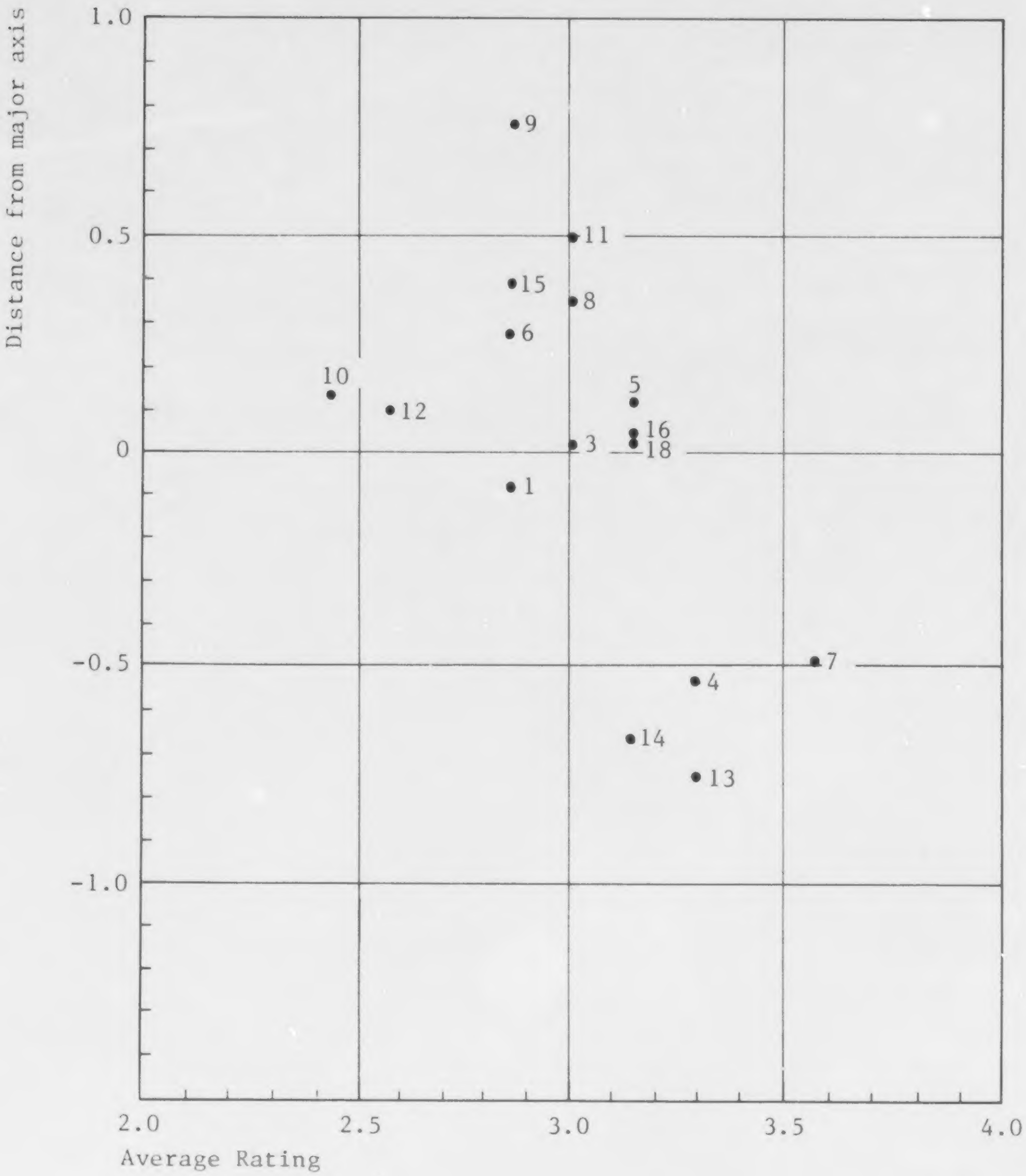


Figure B2. Delta Plot for Verbal Test 1



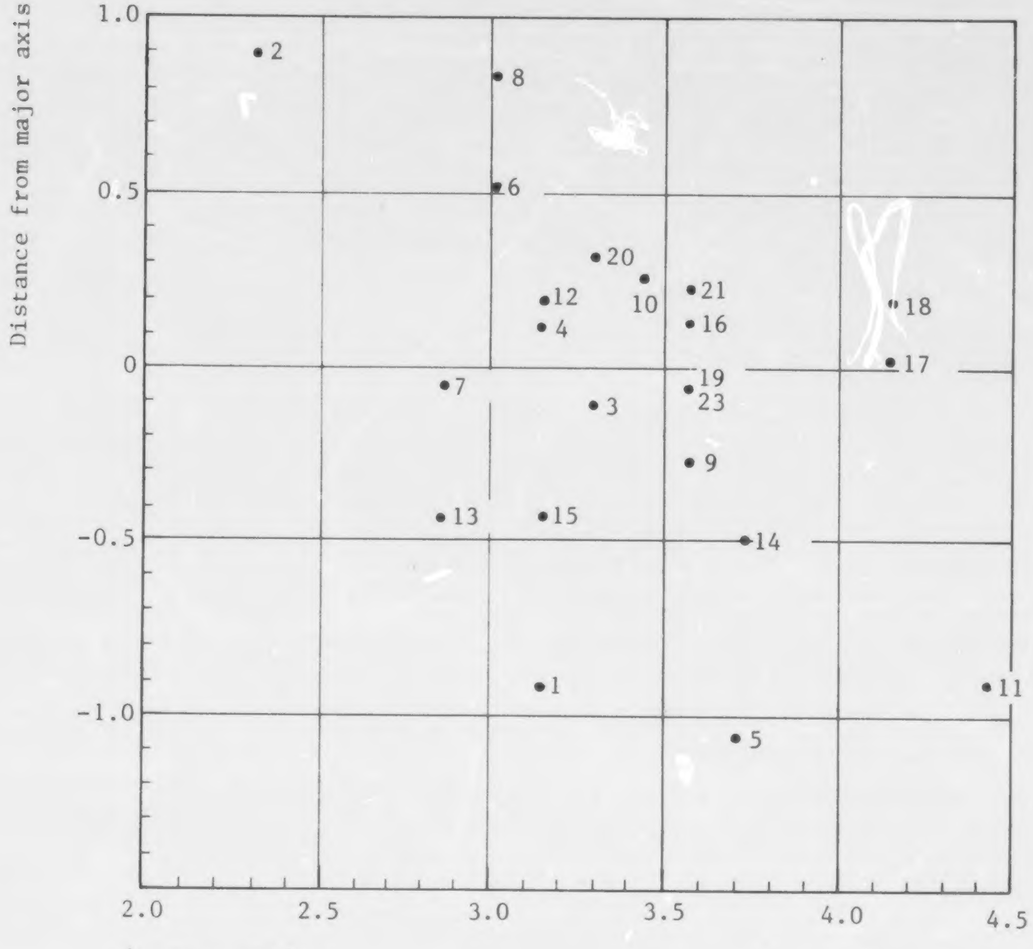
Plot TIS	Test	Form	N	
50015001	SAT-V	LSA33	046	May 1976 Sample
50025002	SAT-V	LSA33	046	May 1963 Sample

Figure B3. Scatter Plot for Math Test 1
Average Rating vs Distance from Major Axis



Numbers refer to the item number in Test 1.

Figure B4. Scatter Plot for Math Test 2
Average Rating vs Distance from Major Axis



Numbers refer to the item number in Test 2.

APPENDIX C

QUESTIONS IDENTIFIED AS EXTREME OUTLIERS

Table C1

The Five Verbal Questions with the Highest Rater Means (High rater mean and negative perpendicular distance both indicate a question that should be easier for the more recent group.)

Question 55, Test 1		
	Rater Mean	3.43
	Perpendicular Distance	-0.20

55. Some firms have so ------- their line of products that, when the seasonal demand for one product slackens, the demand for some other one increases.

(A) consolidated
(B) refined
* (C) diversified
(D) equalized
(E) standardized

Comments on Question 55:

1. Consumer economics vocabulary more in use today in social studies.

---

* Denotes correct answer

Passage for Questions 86-90

The word cybernetics is from the Greek for "steersman," from the same root as the word "governor." It is Wiener's coinage for the new science of "control and communication in the animal and the machine." The element of control is that exemplified in the mechanical device known as a governor, operating on the fundamental principle of the feedback: the working of the machine at a given time is itself a stimulus for the modification of the future course of its working. Communication enters into cybernetics by way of an unorthodox and ingeniously precise concept of "information" as the degree to which the impulses entering a machine reduce the uncertainty among the set of alternatives available to it. Thus if a machine contains a relay which has two possible positions, an impulse which puts the relay into one of these two has conveyed to the machine exactly one "bit" of information. All communication can be regarded as made up of such bits, as is suggested by the dots and dashes of the Morse code, for example. Modern machines actually use a binary arithmetic; i.e., a notation in which all numbers are expressed in terms of powers of 2 (rather than 10) and are therefore written as a series of "zeros" and "ones." One bit of information is conveyed by each choice of either a zero or a one—the word "bit" being an abbreviation for "binary digit."

Working with these concepts of communication and control, cybernetics becomes relevant to the study of man because human behavior is paralleled in many respects by the communication machines. This parallel is no mere metaphor, but consists in a real similarity between the machine processes and those of human behavior. The Darwinian continuity between man and the rest of nature has now been carried to completion: man's rationality marks only a difference of degree from other animals and, fundamentally, no difference at all from the machine. For modern computers are essentially logical machines: they are designed to confront propositions and to draw from them their logical conclusions. With communication and control as the key, a similarity can also be traced between an individual (whether human or mechanical) and a society (again, whether men or machines in a well-designed factory). The metaphors of Plato and Hobbes can now be given a literalist interpretation.

Thus cybernetics bears on the study of human behavior in a variety of ways: most directly, by way of neurology and physiological psychology; and by simple extension, to an improved understanding of functional mental disorders, which Wiener finds to be primarily diseases of memory, thus arriving at Freudian conclusions by a totally different route. Of particular interest are the implications, definitely present though not explicitly drawn, for such classical philosophical puzzles as those concerning free will and the mind-body relation. The analysis of mind and individual personality as a structure of certain information processes renders obsolete not only the "mind substance" of the idealist, but mechanistic materialism as well. Mind is a patterning of information and not spirit, matter, or energy.

Question 86, Test 1

Rater Mean	3.43
Perpendicular Distance	0.08

86. According to the passage, which of the following concepts are fundamental to the science of cybernetics?

(A) Man and machines
*(B) Information and regulation
(C) Computers and logic
(D) Binary digits and bits
(E) Impulses and probabilities

Question 87

Rater Mean	3.71
Perpendicular Distance	-0.49

87. According to the explanation in the passage, which of the following is a binary digit?

(A) $\frac{1}{2}$ *(B) 1 (C) 2 (D) 4 (E) 23

Question 90

Rater Mean	3.43
Perpendicular Distance	-0.49

90. On the basis of the passage, how many "bits" are in 01001?

(A) 2 (B) 3 (C) 4 *(D) 5 (E) 1001

Comments on Questions 86, 87, and 90:

1. Computer related terms are more familiar.
2. Difficult passage in vocabulary, concepts, and demands on analogy.
3. Today's students have more experience with the computer and related vocabulary.
4. Timeliness of content
5. Question 90 is essentially a mathematical question--computers are "in" for many students.

Question 52, Test 2

Rater Mean	3.43
Perpendicular Distance	-0.14

52. Engineers have long ------- that magnetic tape would ------- standard phonograph records, since it wears longer and is difficult to scratch.

*(A) predicted. . replace (B) warned. . destroy
(C) admitted. . ruin (D) lamented. . surpass
(E) charged. . outsell

Comments on Question 52

1. A societal difference--a current technological awareness.
2. Students have much tape and little record experience.

Table C2

The Five Verbal Questions with the Lowest Rater Means (Lower mean indicates the question should be more difficult for the more recent group)

Question 2, Test 1

Rater Mean	2.14
Perpendicular Distance	-0.70

2. Although smallpox has been and still is a severe and dreadful disease, there is some evidence that present-day smallpox has lost part of its -------.

(A) immunity ★(B) virulence (C) effectiveness
(D) symptoms (E) activity

Comments on Question 2:

1. Vocabulary--"virulence" rarely used today--few epidemics written about in news.
2. Difficult vocabulary.

Question 18, Test 1

Rater Mean	2.14
Perpendicular Distance	-0.34

18. EXCULPATE: (A) praise (B) promote
★(C) prove guilty (D) welcome cordially
(E) detract

Comments on Question 18:

1. A rarely used term today.
2. Difficult vocabulary
3. No longer in common usage

Question 58, Test 1

Rater Mean	2.14
Perpendicular Distance	-0.56

58. Like an apéritif, the book stimulates without ------- the -------.

(A) reaching. . climax
(B) calming. . nerves
★(C) satisfying. . appetite
(D) exciting. . emotions
(E) exhausting. . possibilities

Comments on Question 58:

1. Many do not know foreign words.
2. Fewer foreign language electives.
3. Note choice (C).

Passage for questions 76 and 80, Test 1

I feel not in me those sordid and unchristian desires of my profession; I do not secretly implore and wish for plagues, rejoice at famines, consult almanacs in expectation of malignant aspects, fatal conjunctions, and eclipses. I rejoice not at unwholesome springs nor unseasonable winters: my prayer goes with the husbandman's. Let me be sick myself if sometimes the malady of my patient be not a disease unto me. I desire rather to cure his infirmities than my own necessities. I am not only ashamed but heartily sorry that, besides death, there are diseases incurable: yet not for my own sake, or that they be beyond my art, but for the general cause and sake of humanity, whose common cause I apprehend as my own. And to speak more generally, those three noble professions which all civil commonwealths do honor are raised upon the fall of Adam and are not anyway exempt from their infirmities; there are not only diseases incurable in Physick, but cases indissolvable in Laws, vices incorrigible in Divinity. I can cure the gout or stone in some sooner than Divinity pride or avarice in others. I boast nothing, but plainly say we all labor against our own cure; for death is the cure of all diseases. There is no universal remedy I know but this, which, though nauseous to queasy stomachs, yet to prepared appetites is nectar and a pleasant potion of immortality.

76. The author indirectly criticizes some of his colleagues by implying that they

(A) put their faith in universal remedies
(B) are proud and avaricious
*(C) care more for gain than for human welfare
(D) allow superstition to influence their judgments
(E) fail to acknowledge the incurability of some diseases

Rater Mean	2.14
Perpendicular Distance	0.43

---

80. The sentence beginning "I can cure the gout or stone . . ." (lines 22-24) might be clarified by inserting

(A) commas before and after the word "pride"
(B) the words "I can cure" before the word "Divinity"
(C) the words "I can be cured of" before the word "Divinity"
*(D) the words "can cure" after the word "Divinity"
(E) the words "can be cured of" after the word "Divinity"

Rater Mean	2.14
Perpendicular Distance	0.08

Comments on Questions 76 and 80:

1. Close analytic reading of difficult prose has been neglected or abandoned in the many secondary school curriculums thereby making such questions more difficult.
2. Archaic usage and vocabulary
3. Less concern/experience with pre-20th century literature, hence with archaic diction.
4. Task archaic--divine Physick indeed.
5. Obscure construction in question 80.

Table C3

The Five Verbal Questions with the Greatest (Negative) Perpendicular Distance (These questions were relatively easier for more recent groups)

Question 19, Test 1

Rater Mean 2.29
Perpendicular Distance -0.98

19. ESBULLIENCE: *(A) apathy (B) ignominy
(C) antithesis (D) protectiveness
(E) frustration

Comments on Question 19:

1. Difficult vocabulary
2. Uncommon in jargon

Question 21, Test 1

Rater Mean 3.00
Perpendicular Distance -1.05

21. SUFFOCATE: OXYGEN:: (A) electrocute: current
* (B) starve: food (C) write: pencil
(D) burn: paper (E) murder: poison

No comments

Question 27, Test 1

Rater Mean 3.00
Perpendicular Distance -1.12

27. ANALGESIC: PAIN:: (A) submission: authority
*(B) silencer: noise (C) monotony: task
(D) pride: humiliation (E) funeral: death

Comments on Question 27:

1. Formerly analgesic was a common term used in commercials for aspirin--not today.
2. More concern in early 60's with sciences.

Question 66, Test 1

Rater Mean 2.43
Perpendicular Distance -1.19

66. HOYDEN: (A) pleasant companion
(B) generous friend
(C) old woman
(D) aristocratic youth
* (E) well-mannered girl

Comments on Question 66:

1. Archaic terminology
2. Difficult vocabulary
3. "Hoyden" obsolete

Question 12, Test 2

Rater Mean	2.71
Perpendicular Distance	-1.00

12. HOMESTRETCH : RACE :: *(A) finale : opera
(B) goal : contest (C) boundary : journey
(D) terminal : station (E) platform : campaign

No comments

Table C4

The Five Verbal Questions with the Greatest (Positive) Perpendicular Distance
(These questions were relatively more difficult for the more recent group)

Passage for Questions 41–42, Test 1

The ways in which people undertake the waging of war depend not only upon the tools they possess but also upon their social organization and motivations. The chief reasons for war among primitive peoples are not economic. Real or fancied grudges, the desire for revenge, and extroverted antipathy come first. Reciprocity, both positive and negative, is the basis of all social relations. In 1933 an old Comanche gave voice to this in the expressed desire to scalp a young anthropologist "because he looks like a German," and no one had brought old Mumseka a German scalp for the life of his grandson who had fallen in France in 1918.

Religious and magical notions probably rank next after revenge drives. All head-hunting is tied up with supernaturalism and with the belief that dead men's power can be taken with their heads. Polynesian wars were frequently undertaken to obtain sacrificial victims. Ashantis and other West Africans kept prisoners for sacrifice. The bloody religion of the Aztecs called for thousands of war prisoners for human sacrifice to the gods—a fact that proved their undoing when they came to grips with Cortez. The Aztecs fought to take prisoners; the Spaniards fought to kill. They baptized only survivors.

The seeking of status and excitement is another important motivation for war. Pueblo Indians allow no special honor to warriors, but Plains Indians made the measurement of a man his war record. Few are the tribes that do not honor the warrior. Outright conquest was undertaken by Inca, Aztec, and many African kings as a means of enhancing personal glory and honor.

In the period of disillusionment following the First World War, it became the fashion to maintain that all wars spring from sordid economic greed. Self-enrichment is a mighty, driving power. War can serve well as a means of depriving other peoples of access to natural resources. Or it may be used as an instrument of subjugation for the purposes of callous exploitation. Whether it is so used or not depends upon various factors that are culturally determined.

War has become the deadly issue of the twentieth century. It now threatens the very existence of the societies that have brought it to its high peak of destructive perfection. After all, no culture can long nourish a trait that is actively and effectively self-destructive. But we are attempting just this, and it is our tragedy that, seeing where we are heading, we seem unable to alter our course.

41. Which of the following best summarizes and epitomizes the Comanche concept of justice as exemplified by Mumseka?

(A) Do unto others as you would they would do unto you.
(B) Though justice moves slowly, it seldom fails to overtake the wicked.
(C) Let justice be done though the heavens fall.
*(D) An eye for an eye, and a tooth for a tooth.
(E) He injures the good who spares the bad.

Rater Mean	3.14
Perpendicular Distance	1.63

42. In those tribes where warriors were looked upon with favor, what was the principal motivation for individual participation in wars?

(A) To improve the individual's economic situation
(B) To obtain sacrificial victims
*(C) To attain fame and recognition
(D) To propagate the tribal religion
(E) To avenge the tribal honor

Rater Mean	3.29
Perpendicular Distance	1.07

Comments on Questions 41 and 42:

1. If the curriculum change in the past 13 years in social studies has had any effect in making anthropological content and concepts more familiar, today's students SHOULD be more comfortable with this passage.

Passage for question 48, Test 1

The increasingly timid use of the death penalty makes nonsense of the main arguments for retaining it: that it is a necessary form of retribution, an expression of condemnation of a certain type of crime, and a necessary deterrent against this same crime. Clearly, if some first degree murders are punished by death while others are not, society is expressing its condemnation only of particular murderers rather than of first degree murder as such. Hence the fine line between retribution and vengeance disappears.

Historically, capital punishment belongs to a penal system based on violence; the morality which fostered it has long been in retreat before humanitarian and scientific advances. Modern penology seeks to replace the concept of "the punishment to fit the crime" with a new notion: "the treatment to fit the criminal." The interests of society must, of course, be placed before the rights of the individual; no judge should permit dangerous criminals to remain at large. But society's interests would be adequately protected if such criminals were treated before committing a capital crime!

This could happen under a curative penal system. Murder is seldom the first crime committed, and a competent diagnostician usually can detect homicidal tendencies before these erupt. But little improvement can be expected so long as an archaic legal test of sanity—the ability to distinguish between right and wrong—holds abnormal people fully responsible to the law. The practice of punishing these helpless people does more than violate the rights of the individual: it threatens public safety. For, while punishment has no effect on them, its infliction means that society cannot be permanently safeguarded from them until they commit a crime so great that the legal sentence is life imprisonment or death. The law of criminal responsibility must be reformed if the problem that the abnormal offender represents is ever to be solved by curative means. And this is a reform, vital to modern penological principles, that the death penalty and other purely punitive symbols are holding back.

Rater Mean	3.00
Perpendicular Distance	0.79

48. According to the passage, why have the instances of punishment by death become less frequent in recent years?

(A) Laws providing for capital punishment have been modified.
(B) Other forms of punishment have been proven to be more humane.
* (C) Those passing sentence have chosen to give other penalties.
(D) Deterrent influences have made capital crimes less frequent.
(E) Capital punishment has little effect in preventing crime.

Comments on Question 48:

1. In the past 13 years criminology/sociology vocabulary and concepts have become a larger part of the secondary curriculum.
2. Recent concerns about death penalty may cause students to become concerned by opening statement. "Increasingly timid use" suggests there has not been recent publicity about re-instatement of death penalty in California. Some articles see the re-instatement as a backward step, not "increasingly timid use..."

Question 60, Test 1

Rater Mean	2.86
Perpendicular Distance	1.42

60. SECEDE: (A) advance (B) promulgate
*(C) join (D) test (E) veto

Comments on Question 60:

1. Harder because of absence of "civil war" era in conceptual social studies.

Question 75, Test 1

Rater Mean	3.00
Perpendicular Distance	0.93

75. RABBIT : SPEED :: (A) muskrat : water
*(B) bee : industry (C) beaver : dam
(D) fox : evil (E) cat : worship

No comment

Table C5

The Five Mathematical Questions with the Highest Rater Means (High mean indicates question should be easier for the more recent group)

Question 5, Test 2

Rater Mean 3.71
Perpendicular Distance -1.07

5. If $x - 0 = y$, then

(A) $x > y$ *(B) $x = y$ (C) $x < y$

(D) $\frac{y}{x} = 0$ (E) $\frac{x}{y} = 0$

Comments on Question 5:

1. Real number properties emphasized recently at all levels of instruction.
2. Between the two times of administration there was a great deal of emphasis on the properties of 0.

Question 11, Test 2

Rater Mean 4.43
Perpendicular Distance -0.91

11. The symbol $\bullet$ is an operation of arithmetic, a and b are real numbers, and $a \neq 0$. If $a \bullet b = b \bullet a$ and $a \bullet 0 = a$, then operation $\bullet$ is

*(A) addition
(B) subtraction
(C) multiplication
(D) division
(E) either addition or multiplication

Comments on Question 11:

1. 1973 group has had more experience with abstract idea of binary operations.
2. Now included in curriculum.
3. The modern math student would feel comfortable with * operation.
4. Real number properties emphasized recently at all levels of instruction.
5. Direct influence of emphasis in "new Math".

Question 14, Test 2

Rater Mean 3.71
Perpendicular Distance -.50

14. If the product of x and y is divisible by 12 and if y is an odd integer, then

(A) x is divisible by 3
* (B) x is divisible by 4
(C) x is divisible by 6
(D) y is divisible by 6
(E) y is divisible by 2

Comments on Question 14:

1. Number theory used more motivationally in recent years.
2. More work in the curriculum on number theory.

Questions 17-18, Test 2

Questions 17-18 refer to the following definitions.

The notation $x|y$ shall be defined as follows: $x|y = x + y$ if x and y are both odd or both even; $x|y$ = positive difference of x and y if one is odd and one is even. Thus, $1|2 = 1$; $2|4 = 6$; and $3|5 = 8$.

17. $\frac{1|2}{3|4} =$

(A) 0 (B) $\frac{1}{7}$ (C) $\frac{3}{7}$ *(D) 1 (E) 3

Rater Mean 4.14
Perpendicular Distance 0.02

18. If x is positive and $(2x + 1)|(4x + 4) = 9$, then $x =$

(A) $\frac{2}{3}$
(B) 1
(C) 2
(D) $\frac{7}{3}$
*(E) 3

Rater Mean 4.14
Perpendicular Distance 0.19

Comments on Questions 17-18:

1. 1973 group has had more exposure to abstract ideas and binary operations.
2. Definitions of operations are taught today.
3. More emphasis in the curriculum on interpreting notation.

Table C6

The Five Mathematical Questions with the Lowest Rater Means (Low mean indicates question should be more difficult for the more recent group)

Question 10, Test 1

Rater Mean 2.43
Perpendicular Distance 0.13

10. A fisherman caught 10 fish, each weighing 2 pounds, with 8 ounces of bait. At this rate, how many pounds of fish could he expect to catch using 2 pounds of bait?

(A) 20 (B) 40 (C) 60 *(D) 80 (E) 100

Comments on Question 10:

1. Ratio covered at the end of the school year.
2. Less practice in "new math" on these kinds of applications.
3. Go metric.

Question 12, Test 1

Rater Mean 2.57
Perpendicular Distance 0.10

12. If p is the price, in cents, of each apple, how many apples can be purchased for d dollars?

(A) $\frac{d}{p}$ (B) $\frac{p}{100d}$ (C) $\frac{d}{100p}$

(D) $\frac{100p}{d}$ *(E) $\frac{100d}{p}$

Comments on Question 12:

1. Ratio usually covered at the end of the school year.
2. Less practice in the "new math" on these kinds of applications.

Question 17, Test 1

Rater Mean 2.29
Perpendicular Distance -0.78

17. If gasoline costs 28.9 cents a gallon for regular and 31.9 cents a gallon for high-test, what is the difference, in gallons, between the amount of regular and the amount of high-test which $3.00 will buy?

(A) .28 (B) .31 (C) .90 *(D) .98 (E) 3

Comments on Question 17:

1. Problem not closely related to curriculum.
2. This problem would test quite a different skill if the calculator were used. Most students would do this by computing. Where can I buy gas at these prices?

Question 20, Test 1

Rater Mean 2.57
Perpendicular Distance 0.08

20. If 30 per cent of X is 40 per cent of Y, then X is what per cent of Y ?

(A) $33\frac{1}{3}$ (B) 70 (C) 75 (D) 120 ★(E) $133\frac{1}{3}$

Comments on Question 20:

1. Less emphasis on per cent problems.
2. Ratio usually covered at the end of the school year.
3. Seem to do less with per cent.

Question 2, Test 2

Rater Mean 2.29
Perpendicular Distance 0.89

2. The list price of a certain book is $3.95. If it is sold at a 20 per cent discount, how much does the buyer save?

(A) $0.20
★(B) $0.79
(C) $0.95
(D) $1.98
(E) $3.16

Comments on Question 2:

1. Less work on these types of problems.

Table C7

The Five Mathematical Questions with the Greatest (Negative) Perpendicular Distances (These questions were relatively easier for more recent groups)

Question 21, Test 1

Rater Mean 2.86
Perpendicular Distance -1.13

21. An instrument dial is marked off in four numbered points, each 90 degrees apart. The pointer on the dial turns at the rate of 30 degrees per second from 0 to 2. After $2\frac{1}{2}$ seconds, what is the ratio of the distance already covered to the distance remaining to point number 2 ?

(A) $\frac{1}{3}$ (B) $\frac{4}{7}$ *(C) $\frac{5}{7}$ (D) $\frac{2}{3}$ (E) $\frac{5}{6}$

Comments on Question 21:

1. The wording is ambiguous. Supply a picture with the 4 numbered points.

Question 24, Test 1

Rater Mean 3.14
Perpendicular Distance -1.05

24. A model-M truck carries between 4 and 5 tons of bricks and a model-N truck carries between 5 and 7 tons of bricks. What is the smallest possible difference, in tons, between a load on 3 model-M trucks and a load on 2 model-N trucks?

*(A) 0 (B) 1 (C) 2 (D) 3 (E) 5

Comments on Question 24:

1. Inequalities given more study.

Question 1, Test 2

Rater Mean 3.14
Perpendicular Distance -0.92

1. If x times x is 25, then x times x times x is

(A) 25 (B) 30 *(C) 125 (D) 150 (E) 625

No Comments

Question 5, Test 2

Rater Mean 3.71
Perpendicular Distance -1.07

(See Table C5-question also identified by panel as favoring recent group).

Question 11, Test 2

Rater Mean 4.43
Perpendicular Distance -0.91

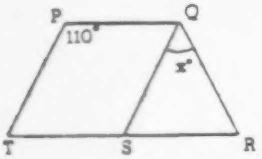
(See Table C5-question also identified by panel as favoring recent group).

Table C8

The Five Mathematical Questions with the Greatest (Positive) Perpendicular Distance (These questions were relatively more difficult for the more recent group).

Question 9, Test 1

Rater Mean 2.86
Perpendicular Distance 0.76



9. In the figure above, PQST is a parallelogram; QR = QS and TR is a straight line. If $\angle P = 110^{\circ}$, then x = (?)

*(A) 40 (B) 50 (C) 70 (D) 110

(E) It cannot be determined from the information given.

No Comments

Question 2, Test 2

Rater Mean 2.29
Perpendicular Distance 0.89

(See Table C6 question also identified by panel as being more difficult for recent group)

Question 8, Test 2

Rater Mean 3.00
Perpendicular Distance 0.83

8. A farmer has completed fencing all but one side of a square field. If he has already used 3n feet of fencing, how many feet will he need for the last side?

(A) $\frac{3n}{4}$ ft. *(B) n ft. (C) $\frac{4n}{3}$ ft.

(D) 3n ft. (E) 4n ft.

Comments on Question 8:

1. Use metric units.

Question 22, Test 2

Rater Mean 3.14
Perpendicular Distance 0.91

22. If $\frac{x}{2}$ is the perimeter of a square, then the area of the square is

*(A) $\frac{x^2}{64}$ (B) $\frac{x^2}{36}$ (C) $\frac{x^2}{16}$ (D) $\frac{x^2}{4}$ (E) $4x^2$

No Comments

Question 1, Test 3

Rater Mean 3.00
Perpendicular Distance 0.69

1. At a certain game, there were 48,000 spectators. If $\frac{1}{3}$ of the spectators bought programs and an average of 200 programs was sold by each program vendor, how many vendors were hired for the game?

(A) 32
(B) 40
*(C) 80
(D) 160
(E) 240

No Comments

Family Configuration Effects and the Decline in College Admissions Test Scores: A Review of the Zajonc Hypothesis

Hunter M. Breland
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

January 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

In three recent publications, Zajonc (1975a, 1975b, 1976) has argued that the mean test score declines in college applicants observed over the past several years are due, in part, to changes in family configurations in the national population. He has predicted that the mean scores will reach a perigee toward the end of the 1970s and then begin to increase sharply in the early 1980s. The purposes of this paper are to review the Zajonc theory, examine its foundations, and attempt to estimate what magnitude of score decline might be a result of family configuration factors.

The Confluence Model

Zajonc and Markus (1975) formulated what they term the confluence model of intellectual development. In this model, the family environment is seen as a causal influence determining much of the intellective ambience of individual experiences. The mathematical formulations of the model predict higher intelligence, generally, for those of smaller families with substantial age spacing between siblings. At another extreme, twins are predicted to be of lower intelligence on the average. (There exists much research evidence to support lower average intelligence for twins.) The confluence model was offered as an explanation of the research findings of Belmont and Marolla (1973) who demonstrated some rather striking effects of birth order and family size on the mean scores of Raven Progressive Matrices (a nonverbal intelligence test). Although the Belmont and Marolla data were from the Netherlands, other large-sample studies have yielded similar results in other countries. Breland (1974) reported remarkably similar patterns of scores on the National Merit Scholarship Qualification Test (NMSQT) in the United States. Record, McKeown, and Edwards (1969) reported almost the same kinds of results in a study of verbal reasoning abilities of eleven-year-old English children. There have been many other studies of the effects of birth order and family size on cognitive abilities, but few with the consistency in findings of these three. Many of the studies have concluded that there are no such effects, and some have observed effects in the opposite direction (i.e., that firstborn are the least intelligent and that members of the larger families are the most intelligent). Schooler (1972) wrote a comprehensive review strongly suggesting that the study of such phenomena was fruitless and should be discontinued.

The confluence model of Zajonc and Markus takes its name from a hypothesized similarity between social processes and fluid flow processes. That is, the social process of the family is seen as resembling the flow of a river in which separate currents make up the river. The currents affect the river, and the river affects the currents. In a similar fashion, the growing individual is influenced by the family environment, and the individual in turn influences the family environment. The process is a dynamic one in which mutually dependent changes in the individual and in the environment occur simultaneously. The confluence model can be applied when a reliable measure is available of the individual quality under examination. Intelligence is one quality for which reliable measures are available. The child entering the world, as a stream enters a river, is confronted by a large number of factors that will have bearing upon his or her intellectual growth. As life progresses, these factors change and the child changes.

In applying the confluence model to the Belmont and Marolla data, Zajonc and Markus limit the environmental factors operating to those represented in the data, viz., birth order and family size. Although the effects of spacing of children can be entered in the model, the Belmont and Marolla data did not include spacing, and accordingly spacing effects could not be checked with the data. There are also other important variables such as the intelligence of parents, parent's absence from the family, and the intelligence of siblings that were not included in the application of the confluence model to the data. Nevertheless, Zajonc and Markus suggest that complex relationships among birth order, family size, and intelligence can be predicted using the confluence model. For example, a firstborn in a family of three would be predicted by the model to achieve a slightly higher level of intelligence than a firstborn who is an only child. This kind of outcome can be predicted because the mathematical relationships used in the confluence model include a consideration of the intellectual value of the teaching role of the firstborn with younger siblings in contrast with the lack of a teaching role for the only child.

Family Configuration and Test Score Decline

With respect to the declines in admissions test scores of college applicants during recent years, the family configuration effects take on importance because of the unusual family configuration patterns emerging as a result of World War II and the considerable change in these patterns that occurred during the years following World War II. Large numbers of new families were started during and immediately following the war, and as a result there were disproportionate numbers of firstborn children in some years. In 1940 over 38 percent of all children born were firstborn, in 1941 over 40 percent were firstborn, and in 1942 almost 43 percent were firstborn. There was a slight reduction in firstborn between 1943 and 1946, but in 1947 the proportion of firstborn again approached 43 percent. By 1950 the proportion of firstborn had decreased to 32 percent, and in 1955 it was less than 29 percent. This decline in the proportion of firstborn children continued until 1962 at which point it began to increase again.

If one assumes that the modal age of persons taking the SAT is 17 years, the peak years of SAT scores (1960-64) can be hypothesized as being related to the peak years of firstborn births (1943-47). And the decline in scores from 1964 to 1976 is concomitant with the decline in firstborn births to 1959. Since the decline in the proportion of firstborn births ceased in 1962, one might predict that the SAT score decline will also cease in 1979. Zajonc has arrived at a similar prediction using mean birth rank instead of proportions of firstborn, since under the confluence model all members of the family play a part in intellectual development. A useful way of thinking about the possible relationship between changing family patterns and the score decline is to imagine the family configurations of persons taking admissions tests in 1963, the year in which the mean scores peaked, in contrast to the probable family configurations of persons taking tests in 1976. In 1963 an unusually high proportion of test takers were probably firstborn because the age group, nationally, was high in proportion of firstborn.[1] Moreover, a high proportion were probably secondborn as well, and secondborn represent another high-scoring group in the post-World War II birth population. In 1947, 71 percent of all births were either firstborn or secondborn, whereas in 1959 only 53 percent of births were of these two types.

Most of the Zajonc work is based upon studies of populations outside the United States and of one population in the United States, the National Merit population of 1965. It is reasonable to ask whether family configuration effects have been observed in SAT populations. Altus, in 1965 and in a number of other papers, reported results of a series of studies conducted at the University of California at Santa Barbara. Table 1 shows the results obtained when students entering in the 1960-61 academic year were studied with respect to the configurations of their families. For both males and females, and in all three family sizes studied, firstborn performed consistently better on the SAT-verbal test. While the same consistency is not evident in the SAT-mathematical means, there is a similar trend in some family sizes.

Table 1. SAT Means by Family Configuration for Students Entering the University of California (Santa Barbara), 1960-1961[a]

Sex	Family size	Birth order	N	Mean SAT-V score	Mean SAT-M score
Male	2	1	174	537.50	560.50
		2	77	527.50	568.75
	3	1	111	532.25	574.00
		2	66	531.50	552.75
		3	30	500.50	553.75
	4	1	32	553.75	539.50
		2	26	511.00	563.00
		3&4	16	547.25	581.75
Female	2	1	328	530.25	492.75
		2	180	510.00	494.50
	3	1	183	526.00	491.75
		2	99	522.50	488.50
		3	56	509.75	485.25
	4	1	70	553.75	511.25
		2	30	516.25	488.75
		3&4	24	488.00	480.75

a. From Altus (1965).

Similar findings were obtained by Lemay (1969) in a study of students entering Oregon State University in the 1963-64 academic year. Lemay's results, shown in Table 2, are more consistent than were those of Altus, and the SAT-mathematical means are seen to follow a step-down progression as do the SAT-verbal means. Although no table is available, Walker and Tahmisian (1967) reported observations at Westmont College like those by Altus and Lemay.

Table 2. SAT Means by Family Configuration for Students Entering Oregon State University 1963-64[a]

Family size	Birth order	N	Mean SAT-V score	Mean SAT-M score
2	1	156	477.45	474.56
	2	145	467.49	463.53
3	1	107	461.75	466.11
	2	103	445.48	452.74
	3	86	450.13	450.65
4	1	50	485.70	473.02
	2	63	452.03	472.58
	3&4	48	442.31	434.23

a. From Lemay (1969).

Estimating the Magnitude of Family Effects

At the extremes, the data of Altus would suggest differences due to family configuration effects of around 40 points on the SAT-verbal scale. Lemay's data show an extreme difference of about the same magnitude on both verbal and mathematical tests. On the surface, therefore, it would appear that the family configuration theory that Zajonc has put forth for the score decline is plausible and could account for practically all the declines that have been observed in the SAT population. Such a conclusion, however, would require that all students taking the SAT in 1963 were the firstborn of two-child families and that all students taking the SAT in 1976 were thirdborn or fourthborn of four-child families (considering the Lemay data of Table 2, for example). In addition, it would be necessary that effects such as those observed by Altus and Lemay in the early 1960s persisted from that time until 1976. And there exists no evidence to confirm that family configuration differences of this type did persist beyond the early 1960s.

To arrive at some kind of rough estimate of the magnitude of possible effects of the type hypothesized by Zajonc, it is useful to assume that the effects persisted over a long period of time. Then one only needs to know something of the changes that have occurred in family sizes and mothering patterns over the period of interest. Evidence of changes in families can be obtained from natality tables of United States vital statistics such as those reproduced in Table 3, which shows birth rates from 1946 to 1967 subclassified by the order of live births. To illustrate, in 1947, there were 47.8 firstborn white births per 1,000 white women between the ages of 15 and 44. In 1960, the birth rate for firstborn was only 30.8. Changes in the birth rates for other birth orders (and family sizes, implicitly) are also shown in Table 3. By comparing the natality data with estimated family configuration effects on SAT scores, assuming that effects observed in only a few years persisted over a long time, it is possible to arrive at an estimate of the magnitude of the possible family configuration effects.

Table 3. U.S. White Birth Rates[a] by Order of Birth and Year from 1946 to 1967[b]

Year	Total Birth Rate	Birth Rate by Birth Order						
		1	2	3	4	5	6&7	8+
1946	100.4	39.5	28.5	14.4	7.3	4.0	3.9	2.8
1947	111.8	47.8	30.8	15.3	7.4	4.0	3.8	2.7
1948	104.3	39.9	31.1	15.7	7.4	3.9	3.7	2.6
1949	103.6	36.3	32.2	16.6	7.9	4.0	3.8	2.7
1950	102.3	33.3	32.3	17.9	8.4	4.1	3.7	2.5
1951	107.7	35.0	32.9	19.5	9.4	4.5	3.9	2.5
1952	110.1	34.1	33.1	21.0	10.4	5.0	4.0	2.5
1953	111.0	33.3	32.9	21.6	11.1	5.4	4.3	2.5
1954	113.6	33.3	32.8	22.6	12.0	5.9	4.6	2.5
1955	113.8	32.6	32.0	22.9	12.6	6.2	4.9	2.5
1956	116.0	33.2	31.9	23.4	13.1	6.6	5.2	2.6
1957	117.7	33.4	31.7	23.7	13.7	7.0	5.6	2.7
1958	114.9	31.9	30.6	23.1	13.8	7.2	5.7	2.7
1959	114.6	31.3	30.0	23.0	14.1	7.4	6.0	2.8
1960	113.2	30.8	29.2	22.7	14.1	7.5	6.1	2.8
1961	112.2	30.7	28.3	22.2	14.0	7.7	6.4	2.9
1962	107.5	29.8	26.9	20.9	13.3	7.5	6.2	2.9
1963	103.7	29.4	25.9	19.6	12.6	7.1	6.1	2.9
1964	99.9	29.8	24.8	18.5	11.7	6.7	5.7	2.7
1965	91.4	28.9	23.0	16.2	10.2	5.8	5.0	2.4
1966	86.4	30.1	22.0	14.4	8.7	4.9	4.3	2.1
1967	83.1	29.7	22.1	13.5	7.9	4.3	3.7	1.8

a. Birth rate is defined as the number of live births per 1,000 white women aged 15-44 years.
b. From Vital Statistics of the United States. Volume 1, Natality. Tables 1-8.

In Table 4, data summarized by Zajonc (1976) are used to make estimates of possible contributions to the SAT score decline of family configuration factors.[2] First, an estimate is made of the incremental value of being of a particular birth order in terms of standard deviations. These estimates are based on four different studies reported

Table 4. Estimate of Magnitude of Family Effects on the SAT Score Decline

Birth order	Estimated intellectual difference (S.D.s)[a]	Estimated SAT-V difference[b]	Proportion in population[c] 1964 (1947)[d]	1976 (1959)[d]	Estimated population change	Estimated contribution to SAT decline
1	+.15	+16.5	.427	.273	-.154	-2.54
2	+.10	+11.0	.275	.262	-.013	- .14
3	-.10	-11.0	.137	.201	+.064	- .70
4	-.20	-22.0	.066	.123	+.057	-1.25
5	-.30	-33.0	.036	.064	+.028	- .92
6&7	-.40	-44.0	.034	.052	+.018	- .79
8+	-.50	-55.0	.024	.024	+.000	- .00
						-6.34

a. Estimated from Figure 1 of Zajonc (1976).
b. Assuming an SAT-V standard deviation of 110.
c. Computed from Table 3.
d. Years in parentheses are the birth years of 17-year-olds.

by Zajonc. Next, assuming an SAT-V standard deviation of 110, it is possible to estimate birth order and family size differences in terms of the SAT scale. These SAT scale differences are then multiplied by changes in the population proportions to determine an estimated contribution of each part of the population change to the SAT score decline. The population proportions shown in Table 4 for 1964 and 1976 were estimated from Table 3 as the ratio of birth rate within birth order to total birth rate. The possible contributions to SAT declines shown in the last column of Table 4 were computed as the difference in population proportions for the two years of interest multiplied by the incremental advantage or disadvantage. Thus, the major portion of the possible decline contribution comes from a decrease in the proportion of firstborn in the population. The total effect on the SAT score decline is estimated at about 6 SAT scale points.[3]

Conclusions

A careful consideration of the hypothesis advanced by Zajonc for part of the SAT score decline has revealed two important results. First, the hypothesis suggests that changing population factors may indeed have been one cause of the SAT score decline. Second, despite the strength of the suggestion, the magnitude of such effects may not be large. Under the assumption that the differences in observed scores categorized by family configuration are due to the environment of the family (Zajonc's theory), the magnitude of the effects is probably not large at all. It is important, however, to emphasize that the possible impact of changing family structures does not depend upon the validity of the confluence model. Regardless of whether the family environment is the cause of the observed differences in mean test scores for different family configurations, the fact remains that such differences have been observed and that the population, nationally, has been changing in its relative proportions of persons with the various family configurations.

Another way of viewing the data that Zajonc has used in developing his theory is that the differences observed are only a manifestation of some other more important factor. The research term used usually in this context is that of a possible intervening variable. It may be that while observed differences in test scores are small when comparing categories of family configurations, they would be large if the intervening variable were known and observable. An anology from Stephens (1968) is useful here. Sex of an automobile driver may be shown to be related to number of accidents, since men have more accidents than women. An important intervening variable in this analogy is number of miles driven, however. Miles driven is associated with number of accidents to a greater degree than is sex. But if miles driven were not known, one would still observe a relationship between sex and number of accidents. It is possible that family configuration and SAT scores are related through some similar unknown intervening variable, and this intervening variable may explain a larger portion of the score decline.

Nonetheless, without some evidence to indicate that there may be an intervening variable of importance, it seems reasonable to conclude that the Zajonc hypothesis could account for only a small portion of the total

SAT score decline. Since the effects observed could have occurred whether or not the confluence model of environmental causation is valid, it seems unnecessary to explore that question.

FOOTNOTES

1. However, many studies have shown that those who attend college are overrepresented by firstborn and other earlyborn.
2. Assuming that distributions by family configuration categories are the same in SAT-taking populations as in national populations. While it is known that such an assumption is not absolutely valid, the differences in the SAT-taking and national distributions may not be large.
3. Schrader (1976) has estimated the effect of the Zajonc hypothesis on the SAT score decline at around 4 or 5 SAT scale points.

REFERENCES

Altus, W.D., "Birth Order and Academic Primogeniture." *Journal of Personality and Social Psychology*, 1965, 2, 872-876

Belmont, L., and Marolla, F.A., "Birth Order, Family Size, and Intelligence." *Science*, 1973, 182, 1096-1101

Breland, H.M., "Birth Order, Family Configuration, and Verbal Achievement." *Child Development*, 1974, 45, 1011-1019

Lemay, M., Birth Order and Scholastic Aptitude and Achievement. Research Bulletin, Center for Research on Student Life and Development, Oregon State University, 1969. Mimeographed

Record, R.G., McKeown, T., and Edwards, J.H.,"The Relation of Measured Intelligence to Birth Order and Maternal Age." *Annals of Human Genetics*, 1969, 33, 61-69

Schrader, W.B., Memorandum. Princeton, N.J.: Educational Testing Service, May 27, 1976

Schooler, C., "Birth-Order Effects: Not Here, Not Now!" *Psychological* Bulletin, 1972, 78, 161-175

Stephens, W.N., *Hypotheses and Evidence*. N.Y.: Thomas Y. Crowell Co., 1968, 204

U.S. Government Printing Office, *Vital Statistics of the United States*. *Volume 1, Natality*. Washington, D.C.: U.S. Government Printing Office, 1967, Tables 1-8

Walker, C.E., and Tahmisian, J., "Birth Order and Student Characteristics: A Replication." *Journal of Consulting Psychology*, 1967, 31, 219

Zajonc, R.B., "Birth Order and Intelligence: Dumber by the Dozen. *Psychology Today*, 8 (8), January 1975, 37-43 (a)

Zajonc, R.B., "Intellectual environment and intellectual development: New data and new implications." Invited paper prepared for the Forty-Seventh Annual Meeting of the Midwestern Psychological Association, May 2, 1975 (b)

Zajonc, R.B., "Family Configuration and Intelligence." *Science*, April 16, 1976, 19, 227-236

Zajonc, R.B., and Markus, G.B., "Birth Order and Intellectual Development." *Psychological Review*, 1975, 82, (1), 74-88

The SAT Score Decline: A Summary of Related Research

Hunter M. Breland
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

January 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $3 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

CONTENTS

Introduction 3
The SAT 6
  Difficulty of the SAT 6
  Relevance of the SAT and Curricular Change 7
The SAT Population 8
  SAT User Colleges 9
  SAT Repeaters 9
  SAT Score Distributions 11
The College-Bound Population 14
  Trends in ACT Scores 14
  College Attendance Patterns 16
  Minorities 19
  Low-SES Students 22
  Low-ability Students 22
The Schools 22
  Proportions of Cohorts Graduating 24
  Abilities of High School Graduates 24
Societal Factors 27
Conclusions 29
Footnotes. 29

TABLES AND FIGURES

Tables

1. SAT-Verbal and SAT-Mathematical Means: Two Methods of Computation 5
2. Comparison of Identical SAT Sections for Two Time Intervals 7
3. Number of SAT Scores Received by Colleges and Universities Grouped by Carnegie Commission Classification in 1960-61, 1966-67, and 1973-74. 10
4. Mean Gains in SAT-Verbal and Mathematical Scores from Junior to Senior Year 12
5. Percentages of SAT Scores in Given Score Ranges 13
6. Means and Standard Deviations for College-Bound Students Taking the ACT Tests in Successive Years 15
7. College Enrollments During the Period of the SAT Score Decline as a Function of the High School Cohort and Institution Type 18
8. College Enrollments of Minorities 21
9. Percentages of High School Graduates Attending Two- or Four-Year Colleges 23
10. Average PSAT Scores for High-School Juniors from Three Norming Samples 25
11. Changes in Grade Equivalents for Grade 8 Students on Selected ITBS Tests 26

12. Mean Reading Scores of 11- and 15-year-Old English Children Since 1948 28

Figures

1. SAT-Verbal and SAT-Mathematical Means for the Period 1957-1975 4
2. Numbers of High School Graduates and Immediate College Entrants 17
3. Rates of High School Graduation and Immediate College Entry for the 18-Year-Old Cohort 20

INTRODUCTION

Since 1963, SAT score averages have declined each year for 12 successive years. Verbal score averages have declined a total of more than 40 points, and a similar but less dramatic decline has occurred in the mean SAT-mathematical score. The nature of these declines is depicted in Figure 1. In Table 1, the values shown in Figure 1 are contrasted with other interpretations of the declines based upon different ways of computing the yearly averages. During the past two or three years numerous explanations have been offered for these declines. It is the intent of this paper to present the rationale of some of the more prominent explanations and to cite available research evidence relating to each.

All the explanations relate to some kind of change that has occurred concurrently with the score declines, and these explanations fall into four general areas of classification: the test, the population taking the test, the schools, and societal factors. Explanations suggesting that the problem resides within the test, although less common than other explanations, usually point to the possibility that the test may have become more difficult. It is also sometimes suggested that tests like the SAT are simply irrelevant in a modern world where written materials are rapidly being replaced with other communication media.

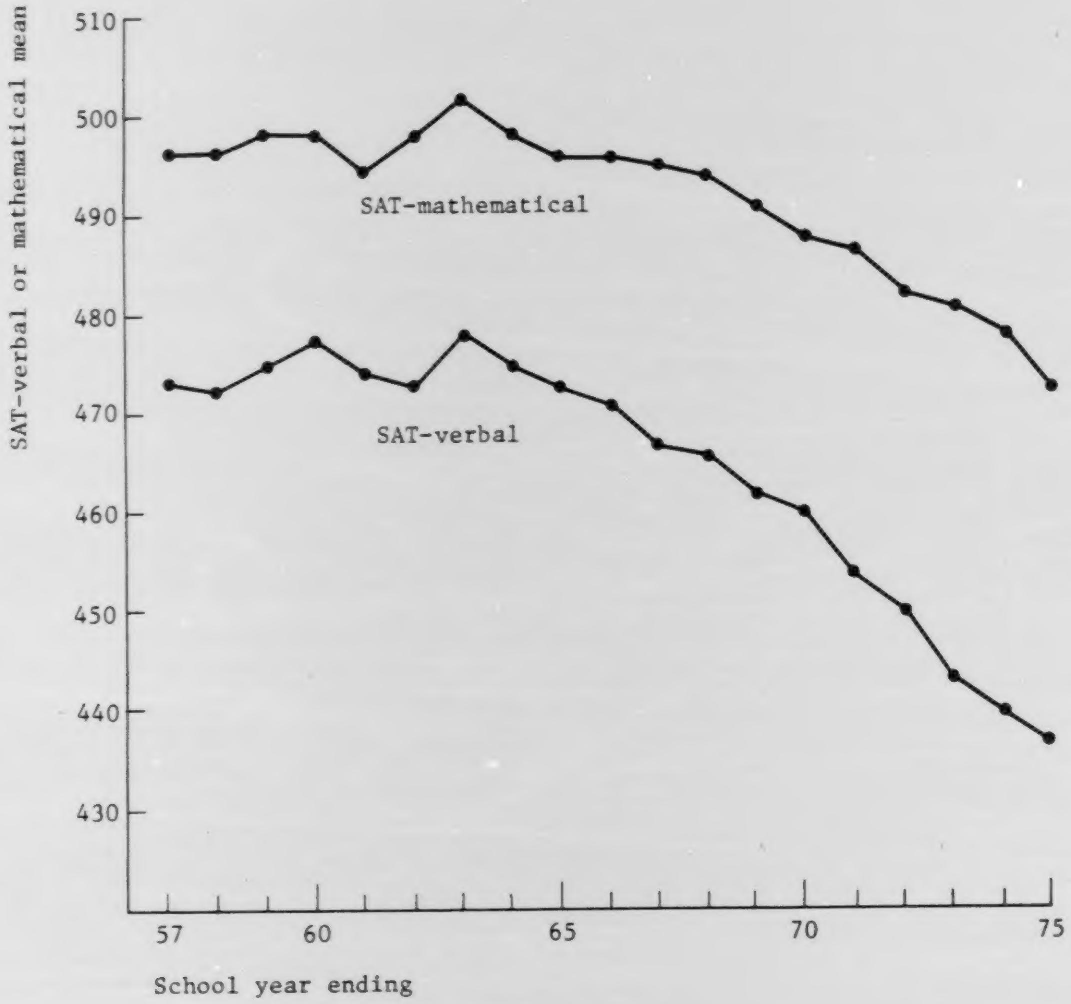
Others argue that, because of emerging egalitarian philosophies--coupled with financial difficulties in postsecondary institutions--many colleges are no longer selective. As a result, students may have been admitted to colleges in the past decade who would have been rejected in previous years. A symptom of this trend in college admissions, it is alleged, is a trend toward increasingly easier college textbooks, since it is claimed that today's college students cannot read as well as did students of earlier years. Of course, the population of students taking the SAT may not be representative of the college-bound population, and changes in the SAT candidate population over the period of the declines may have been different from changes occurring in the total college-bound population.

The most popular set of explanations shifts the burden from the colleges to the schools. In particular, there is speculation about the quality of secondary education in America. In their attempts to be progressive, it is said, schools have become too permissive, and the result has been that traditional academic skills are no longer learned. Those who espouse this explanation often note the increasing constraints on the teaching profession, the encroachment of the courts, and reduced expenditures for education.

A final set of explanations focuses on a broad range of societal factors. Advocates of these explanations note trends among adolescents toward drug use, alcoholism, television watching, and other similar factors that would appear to have detrimental effects on the development of academic aptitudes. Additionally, some would relate the decreased competition for college admission and the use of nonacademic criteria in awarding scholarships to a reduction in the motivation to perform well on admissions tests.

In the sections that follow, explanations such as these are examined in the context of research evidence. In recognition of the responsi-

Figure 1. SAT-Verbal and SAT-Mathematical Means for the Period 1957-1975*



*These means represent all scores reported. Thus, persons having more than one score are counted more than once.

bility of the test developer, the order of consideration begins with the SAT itself. From a research perspective, the remaining explanations are best examined with respect to evidence pertaining to specific populations. These remaining explanations are considered in an order of increasing population size. Thus, following the discussion of the SAT, first the SAT population is examined, next the college-bound population, next the population of high school graduates, and finally extranational populations.

Table 1. SAT-Verbal and SAT-Mathematical Means: Two Methods of Computation

School year	All scores[a] V	All scores[a] M	College-bound[b] seniors only V	College-bound[b] seniors only M
1956-57	473	496	---	---
1957-58	472	496	---	---
1958-59	475	498	---	---
1959-60	477	498	---	---
1960-61	474	495	---	---
1961-62	473	498	---	---
1962-63	478	502	---	---
1963-64	475	498	---	---
1964-65	473	496	---	---
1965-66	471	496	---	---
1966-67	467	495	466	492
1967-68	466	494	466	492
1968-69	462	491	463	493
1969-70	460	488	460	488
1970-71	454	487	455	488
1971-72	450	482	453	484
1972-73	443	481	445	481
1973-74	440	478	444	480
1974-75	437	473	434	472

a. These are the values presented in Figure 1. These means represent all scores reported. Students represented in these means are aggregated without regard to their level of preparation, and students are counted as many times as they have taken the SAT.

b. From College-Bound Seniors, 1974-75. In these means each student is counted only once. For students who have taken the SAT more than once, only the most recent scores earned prior to graduation are counted. The SAT means for the 1967-71 high school classes are estimated values. Corresponding estimates for high school classes prior to 1967 are not available.

THE SAT

Although the instrument itself has not normally been the focus of explanations, it is appropriate to ask whether some change in it might have contributed to the score decline. If the items that make up the SAT now require more complex mental processes than they originally did, a decline in average performance might result. If the SAT items contain material no longer normally learned, either at home or in school, it would be expected that recent cohorts would perform poorly. Or, it could be, in an advanced technological society dominated by television and other nonwritten forms of communication, skills such as those examined in the SAT are no longer necessary or relevant.

Difficulty of the SAT

Since the SAT is changed each year as different forms are generated, and since it has changed over the years to make it more responsive to the needs of the educational community and to improve individual questions where possible, there has always been the concern that the test could become either slightly more difficult or slightly easier as such changes are introduced. Over a number of years, the small changes in difficulty could grow systematically and result in a substantial drift in the score scales. To control for this kind of drift, a statistical procedure is regularly followed to insure that reported scores are directly comparable from one form of the test to any other form. This control of difficulty is achieved by retaining identical items in test forms in a systematic way so that all forms can be reported on the same score scale. Test-form "equating" of this kind has been a paramount concern in all College Board test forms since 1941. The equating is done to control for normal form-to-form variations in difficulty. Indeed, were it not for these procedures, discussions of score decline would be almost meaningless. A detailed description of the SAT equating procedures is given in Angoff (1971a, 1971b). The difficulty of maintaining perfect scale stability should be recognized. With respect to this particular problem of scale stability, Wilks (1961) observed:

"It may not be possible to get a definitive answer to this question even under the most ideal conditions of investigation since invariance depends not only on such factors as the reliability of the tests, the effectiveness of statistical techniques of equating and sampling fluctuations in groups of candidates used for equating, but also on the invariance of the performance of succeeding generations of candidates on the subject matter used for equating the successive forms." (p. 15)

Periodically, studies are conducted to check the equating mechanisms. One equating study, Stewart (1966), showed that the SAT scale had been stable during the period 1944 to 1963. The most recent study of this type, Modu and Stern (1975), examined the period 1963-1973, a period during which SAT mean scores have declined substantially. The results obtained by Modu and Stern suggested that during the period of the score decline some drift in the SAT scale did, in fact, occur. However, the scale drift indicated by the study was such as to cause recent scores to be higher than they should have been. Accordingly, a score decline would not appear to be attributable to the scale drift. Since difficulties in sampling and in other procedures may have affected the precision of the Modu and Stern analysis, a study designed to determine more precisely the magnitude of the long-term drift in the SAT scale is currently underway (Marco and Stewart, 1975).

A clearer picture of the difficulty issue is provided by the performance of samples on test sections used for equating--that is, sections that are identical across test administrations and testing years. During the past decade, new samples of the SAT population have consistently scored lower on the same material than old samples. Data from the Modu and Stern study, presented in Table 2, illustrate this trend.

In the case of the first verbal comparison in Table 2, a reprinted section of a verbal test given in 1963 (Form bm) was readministered in 1973--and in the same month, December.[1] Out of the 40 items in this reprinted section, the 1963 sample averaged in excess of 18 correct; the 1973 sample averaged less than 14 correct. For the shorter period from 1966 to 1973, the verbal decline is slightly less than that from 1963 to 1973, as would be expected. The same pattern is evident in the comparisons of the mathematical sections in Table 2. In 1963, the sample obtained an average of more than 10 of 25 mathematical items correct. In 1973, when the same 25-item section was reprinted and administered, the sample obtained less than an average of 8 items correct. And the mathematical decrease from 1966 to 1973 is slightly less than the mathematical decrease for the longer period from 1963 to 1973.

Relevance of the SAT and Curricular Change

Although the skills required by the SAT are in many cases similar to skills acquired in schools, it should be emphasized that the test was not designed to measure school performance. The SAT was designed

Table 2. Comparison of Identical SAT Sections for Two Time Intervals

Test-section type (form)	Number of items	When administered	Sample size	Mean
Verbal (bm)	40	Dec. 1963	1,500	18.43
Verbal (bm)	40	Dec. 1973	5,000	13.99
Verbal (bh)	40	Dec. 1966	5,000	17.23
Verbal (bh)	40	Dec. 1973	4,915	13.99
Mathematical (bk)	25	Dec. 1963	1,500	10.41
Mathematical (bk)	25	Dec. 1973	5,000	7.90
Mathematical (bg)	25	Dec. 1966	5,000	10.94
Mathematical (bg)	25	Dec. 1973	4,921	8.91

to measure long-term developed verbal and mathematical reasoning abilities--attributes that are related to college performance. Therefore, the relevance of the SAT is best evaluated by its predictive validity. Numerous research studies have demonstrated the predictive validity of the SAT during the period of the score decline (e.g., Cleary, 1968; Temp, 1971; Fincher, 1974).

In a culture dominated by television, it is sometimes suggested that the reading and other fundamental skills measured by the SAT are less important than they once were. The increased use of audiovisual aids and materials in the schools may even further intensify the oral-aural influence of television. In recent years the schools have also emphasized other nontraditional kinds of learning, such as noncognitive, interpersonal, and career skills. With these kinds of changes in the schools and in society, some believe that the questions used in the SAT may be more difficult for recent high school populations even though their abilities are as great or greater than ever. Nonetheless, while the schools may have changed, the relationship between SAT scores and college performance has not.

Were the SAT decline only in terms of mathematical skills, a curricular explanation might be available. Recent high school students who grew up with the "new math" would be expected to perform less well on those portions of the mathematics test that deal with the more traditional arithmetic and algebraic skills. As a result, the average SAT-mathematical scores of recent students would be lower. However, this kind of curricular change does not explain the more pronounced decline in SAT-verbal score averages. Moreover, curricular explanations fail to consider the probability that SAT scores are substantially influenced by experiences outside school, including family life and other factors.

Accordingly, the relevance of the SAT cannot be assessed in terms of its similarity to school curriculums. Nor should the SAT be judged entirely on the basis of its demonstrated use in predicting college performance. With open admissions policies and the proliferation of community colleges, one of the more important uses of the SAT today is for student placement in courses of instruction. Additionally, SAT scores in the aggregate are useful to institutions as management information and to the larger society as an indicator of the nature of the college-bound population. Finally, SAT scores tell the student something about himself. Although the language of the test is standard English, this is also the language of most books, journals, newspapers, and magazines--and it is the language of upward mobility in American society.

THE SAT POPULATION

If the SAT itself has not changed sufficiently to account for the decline in mean scores, then the next logical assumption is that the population of students who take the test has changed. Changes in the population taking the test could have resulted from a number of factors. Principally, there may have been a change in the mix of colleges requiring the SAT, changes in the way students repeat or do not repeat the test, changes in the gains they make when repeating, and changes in the attitudes of students toward the admission process--or changes of all three kinds in varying combinations.

SAT User Colleges

When the College Board was first formed, it consisted primarily of prestigious Eastern colleges, but over the years this situation has changed. And, in recent years, some institutions have changed to an "open admissions" policy and have discontinued membership in the Board. At the same time, large numbers of community colleges have been founded, some of which require their students to take the SAT. This change in College Board colleges, and thus perhaps in the kinds of students served, may have contributed to the decline in the SAT mean scores. Although data on the composition of the candidate group over the period of the score decline are very limited, annual summaries have been made of the number of score reports sent to specific colleges. A study currently in progress (Schrader) focuses on the use of those score-report data to explore changes in the College Board candidate population. In this study, tabulations of SAT score reports received by institutions are being made by state, by Carnegie Commission institution type,[2] and by institutional requirements for SAT or American College Testing Program (ACT) tests. Preliminary results from this investigation are presented in Table 3. This preliminary analysis suggests noticeable changes in the patterns of score reporting for the academic years 1960-61, 1966-67, and 1973-74. For example, from 1960-61 to 1973-74, the proportion of SAT score reports going to the most selective liberal arts colleges decreased from 13.2 percent to 5.6 percent, and the proportion of score reports going to two-year colleges increased from 1.9 percent to 8.6 percent. These results make it clear that College Board candidates have changed substantially in their patterns of college applications during the period studied. The extent (if any) to which these changes contributed to the score decline, however, cannot be estimated from this study.

SAT Repeaters

A subset of students who take the SAT repeat the test one or more times. At least two different kinds of score-decline explanations have been suggested that relate to repeaters. First, it has been noticed that in recent years a smaller proportion of students who take the SAT as juniors repeat it as seniors and that a steady decline in this proportion has occurred since about 1962. Since this decline in repeating proportion covers about the same period of time as the score decline, it is attractive as a possible explanation. Nevertheless, both Marco and Stern (1974) and McCandless (1975), after analyses of this kind of junior-senior repeater hypothesis, concluded that no more than about 3 points of the score decline could be accounted for in this way.

A second kind of repeater hypothesis holds that students who repeat the test today are gaining less in their scores than those who repeated several years ago. That is, it is possible that there has been a

Table 3. Number of SAT Scores Received by Colleges and Universities Grouped by Carnegie Commission Classification in 1960-61, 1966-67, and 1973-74[a]

		Number of SAT reports (in thousands)			Percent of SAT reports		
Classification	Control	1960-61	1966-67	1973-74	1960-61	1966-67	1973-74
Research universities I[b]	Public	189	347	345	13.0	12.1	13.0
	Private	152	192	171	10.5	6.7	6.4
Research universities II[b]	Public	151	286	199	10.4	10.0	7.5
	Private	73	96	65	5.0	3.3	2.5
Doctoral-granting universities[b]	Public	75	221	245	5.2	7.7	9.2
	Private	112	166	122	7.8	5.8	4.6
Comprehensive universities and colleges	Public	141	526	502	9.7	18.4	18.9
	Private	177	313	244	12.2	10.9	9.2
Liberal arts colleges I[c]		191	224	150	13.2	7.8	5.6
Liberal arts colleges II[c]		108	260	209	7.5	9.1	7.8
Other specialized (including service academies)		34	33	33	2.3	1.1	1.3
Two-year colleges		28	141	229	1.9	4.9	8.6
Other institutions (including unclassified and unknown)[d]		19	59	145	1.3	2.1	5.5
Total		1,449	2,864	2,660	100.0	100.0	100.0

Footnotes for Table 3

a. Preliminary. From Schrader (1975).
b. The classification of doctoral-granting institutions was based on amount of federal financial support and on number of Ph.D.s (and M.D.s) granted. Research universities I ranked highest on these bases, research universities II second, and doctoral-granting universities third.
c. The classification of liberal arts colleges was based on Astin's selectivity index and on the number of graduates who earned doctorates. Liberal arts colleges I ranked higher on these bases.
d. This group included a variety of professional and specialized universities not classified in the Carnegie Commission report in any of the categories shown above and also included institutions (e.g., hospital schools of nursing, two-year colleges) not listed in the report.

---

reduction in the gain achieved by those who do repeat and that this reduction could have resulted in a decline in the overall mean. Table 4 shows the mean gains in SAT scores for three points during the years of the score decline for different repeating patterns. At the bottom of Table 4, weighted average gains are presented. These weighted average gains show an overall reduced gain of about 8 points on SAT-verbal and about 9 points on SAT-mathematical. For the intermediate point (1967-68), the reduction in gain is more. Since only 200,000-300,000 of the more than one million students taking the test are represented in these junior-senior repeaters, however, little impact on the overall SAT means could be accounted for by these reduced gains shown. Of course, there are others who repeat in addition to those who repeat in the junior to senior pattern--including those who repeat during their senior year and those who repeat after graduation. Additionally, there exists a large group of "repeaters" who take the PSAT/NMSQT[3] (which is similar to the SAT) as juniors and also take the SAT as seniors.

SAT Score Distributions

Changes in SAT score distributions during the period of declines in mean scores are shown in Table 5. The percentage of students scoring in the 700-800 range on the verbal sections of the SAT declined from 2.3 percent in 1967 to 1.1 percent in 1975. In the 600-700 range, the percentages dropped from 10.2 percent in 1967 to 7.0 percent in 1975. A similar but smaller drop in the high scores occurred for the mathematical sections of the SAT.

These percentage reductions in the high score ranges could have resulted merely from greater numbers in the lower ranges, and such a trend seems reasonable in view of apparent increases in students of low socioeconomic status (SES) entering college in recent years. There is, however, an anomaly that is as yet unexplained. The absolute number of

Table 4. Mean Gains in SAT-Verbal and SAT-Mathematics Scores From Junior to Senior Year

Repeat pattern		Testing year of final score		
		1962-1963	1967-1968	1973-1974
March to November	V	---	---	17.3
	M	---	---	21.3
	N	---	---	25,758
March to December	V	42.1	19.9	19.4
	M	33.2	18.0	17.9
	N	44,930	67,153	10,923
April/May to November	V	---	---	23.4
	M	---	---	21.0
	N	---	---	110,996
April/May to December	V	24.1	14.0	24.4
	M	23.3	13.5	15.4
	N	96,474	207,094	43,691
July to November.	V	---	---	14.9
	M	---	---	4.8
	N	---	---	24,618
March to Jan./Feb.	V	40.1	25.0	17.8
	M	38.0	19.0	25.9
	N	5,891	3,938	1,021
April/May to Jan./Feb.	V	22.9	19.3	22.0
	M	28.8	14.0	20.4
	N	16,634	13,008	4,580
July to Jan./Feb.	V	---	---	17.4
	M	---	---	13.9
	N	---	---	1,342
Weighted average.	V	29.5	15.8	21.2
	M	27.1	14.6	17.8
	N	163,929	291,193	222,929

Table 5. Percentage of SAT Scores[a] in Given Score Ranges

Section	1966-1967	1967-1968	1968-1969	1969-1970	1970-1971	1971-1972	1972-1973	1973-1974	1974-1975
Verbal									
700-800	2.3	2.0	1.8	1.7	1.7	1.4	1.1	1.2	1.1
600-699	10.2	10.1	9.9	10.1	9.4	9.0	8.0	7.7	7.0
500-599	23.9	23.4	23.9	23.6	23.5	22.8	21.5	20.6	21.1
400-499	35.4	34.6	33.6	33.1	33.8	33.8	33.6	34.2	34.2
300-399	22.9	24.5	23.8	25.0	25.1	25.9	27.9	28.0	27.4
200-299	5.3	5.4	6.0	6.6	6.4	7.2	7.8	8.2	9.2
Mean	467	466	462	460	454	450	443	440	437
SD	109	108	108	110	109	108	107	107	108
Math									
700-800	3.9	3.6	3.5	3.5	3.6	3.0	3.2	3.4	3.3
600-699	14.0	14.6	14.1	14.0	14.5	13.9	13.2	13.0	12.3
500-599	29.1	28.6	28.6	28.5	27.9	28.4	28.8	27.0	26.3
400-499	32.7	31.0	31.2	31.0	31.4	30.8	30.7	30.5	30.1
300-399	17.4	18.3	18.7	18.9	19.0	19.9	20.1	21.5	23.0
200-299	2.9	3.8	3.8	4.1	3.6	3.9	4.1	4.6	5.0
Mean	495	494	491	488	487	482	481	478	473
SD	110	112	111	113	113	112	112	115	114

a. These scores represent all scores reported. Thus, persons having more than one score are counted more than once.

students scoring in the very high ranges of SAT scores has also declined. In 1967, more than 30,000 persons scored in the 700-800 range on the verbal sections of the SAT; this absolute number of scores in the 700-800 range declined gradually until, in 1975, it was only about 15,000. This gradual decrease in scores in the high ranges has occurred despite fluctuations in the total numbers of students taking the SAT.

An attitudinal hypothesis is one of the few that would explain this drop in scores in the high ranges. Those who repeat, and others, may have changed in their attitude concerning the SAT. For example, one can imagine that, during periods when college access is highly competitive, students may strive for high scores either the first time the test is taken or by repeating the test one or more times. A change in attitude would follow from the situational context of recent years, during which colleges have actively recruited students rather than carefully selecting from large applicant pools.

THE COLLEGE-BOUND POPULATION

To a degree, the SAT population represents the college-bound population, but it must be recognized that not all persons seeking college admission take the SAT. Many institutions require the test administered by the American College Testing Program (the ACT), other institutions allow an option of the SAT or the ACT, and some institutions require neither. Most college-bound students, however, take either the SAT or the ACT, and some take both. In examining the college-bound population, therefore, the experiences of the ACT are of special interest, since if both testing programs have had similar experiences, those experiences would be representative of the college-bound population. On the other hand, if these two national testing programs have had different experiences, the causes would seem to reside within the programs and not in the college-bound population.

Trends in ACT Scores

Consider the decline in mean ACT scores shown in Table 6. Between 1965 and 1974, the mean ACT English score declined from 18.7 to 17.6 (1.1 ACT scale points), the mean ACT Mathematics score from 19.6 to 18.1 (1.5 ACT scale points), and the mean ACT Social Studies score from 20.6 to 17.9 (2.7 ACT scale points). The fourth part of the ACT test, Natural Science Reading, did not show a significant change in mean score. The four ACT subscores are combined to form what is known as the ACT Composite score. For 1975, the ACT Composite score mean was reported to be 18.6--a decline of 1.3 ACT scale points since 1965, when the ACT Composite score mean was 19.9.

If one assumes that the SAT standard deviation is constant at about 110 and that the ACT standard deviation is constant at about 5.5, the SAT "composite" score (combining verbal and mathematical) decline between 1965 and 1975 is about 31/110=.28, and ACT Composite mean score

Table 6. Means and Standard Deviations for College-Bound Students Taking the ACT Tests in Successive Years[a]

Year	English		Mathematics		Social Studies		Natural Science		Composite	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1964-65	18.7	5.1	19.6	6.6	20.6	6.4	20.4	6.1	19.9	5.2
1965-66	19.1	5.1	19.5	6.7	20.5	6.4	20.5	6.1	20.0	5.2
1966-67	18.5	5.3	18.7	7.0	19.6	6.6	20.1	6.3	19.4	5.4
1967-68	18.1	5.3	18.3	7.3	19.4	6.7	19.8	6.5	19.0	5.5
1968-69	18.4	5.2	19.2	6.9	19.4	6.7	20.0	6.4	19.4	5.3
1969-70	18.1	5.3	19.5	6.7	19.3	6.8	20.5	6.1	19.5	5.3
1970-71	17.7	5.6	18.7	7.2	18.3	7.2	20.2	6.4	18.9	5.6
1971-72	17.6	5.6	18.6	7.3	18.4	7.3	20.3	6.5	18.8	5.7
1972-73	17.8	5.4	18.8	7.2	18.1	7.5	20.5	6.5	18.9	5.8
1973-74	17.6	5.4	18.1	7.5	17.9	7.6	20.6	6.5	18.7	5.8
1974-75									18.6[b]	

a. From Ferguson and Maxey (1975).
b. From ACTivity, October 1975.

decline is about 1.3/5.5=.24. The overall declines in SAT and ACT means, therefore, would appear to be of the same order of magnitude.[4] The SAT-verbal score means and the ACT Social Science reading score declined more than did the overall means.

Since almost all the students in the college-bound population take either the SAT, the ACT, or both, these declines are for the most part representative of the college-bound population.

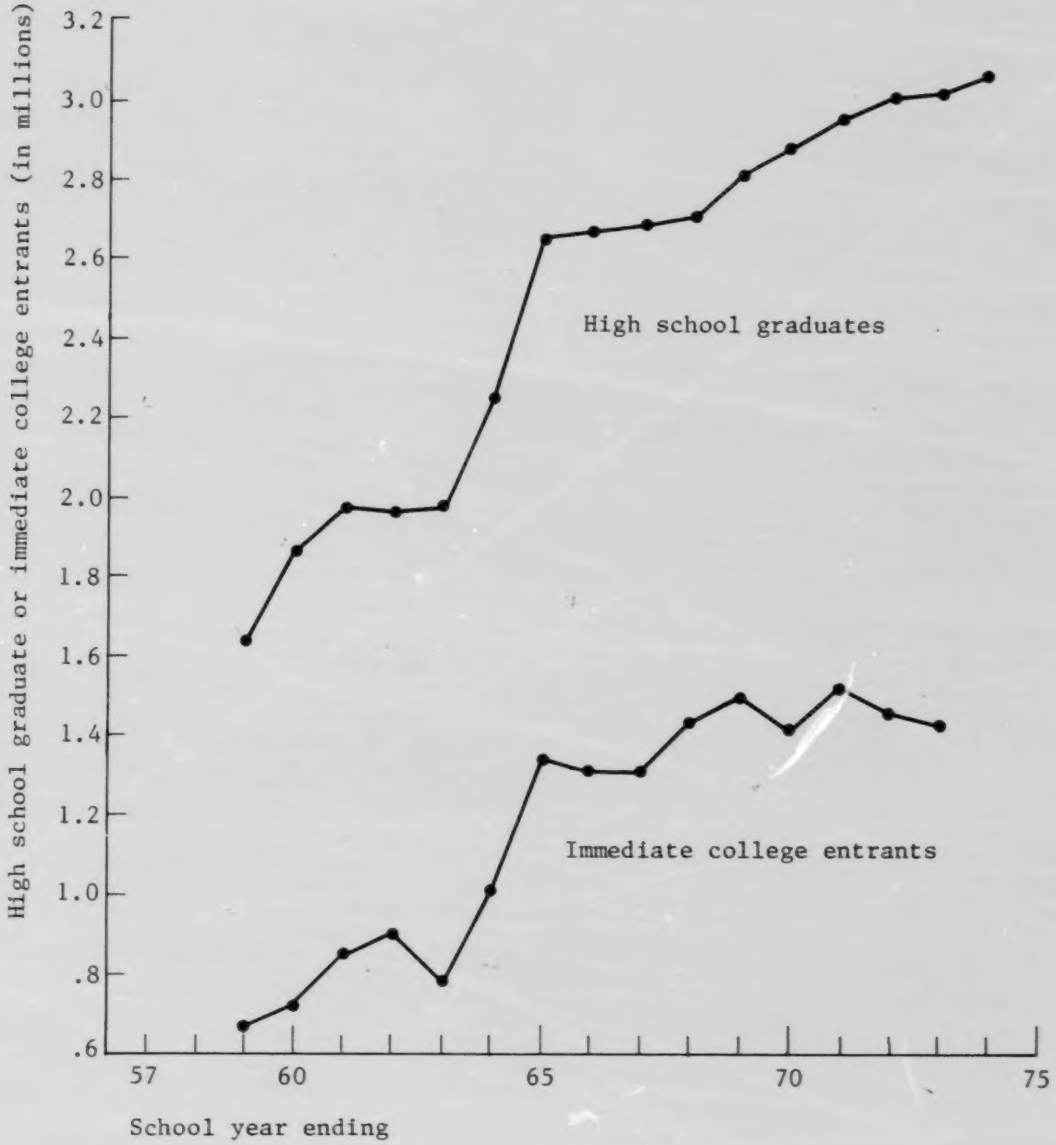
College Attendance Patterns

To gain an initial sense of the relationship between these SAT and ACT score declines and college-going, it is useful to consider the score decline patterns with the college attendance patterns. Figure 2 shows the numbers of high school graduates and immediate college entrants during the period of the score declines. Notice that the year in which the score declines began, 1963, is the same year that the numbers of immediate college entrants began to increase markedly. The increase in college-going reflects, of course, the post-World War II population boom that began in 1946. On the surface, however, there would seem to be no particular reason why an increase in the numbers of persons going to college should cause the academic abilities of those persons to decline. College enrollment could increase simply as a result of a population increase in the college-bound age group, because greater proportions of this age group are considering college, or because of combinations of these population and choice factors. But choice is influenced by high school graduation, since few who do not graduate from high school have the option of attending college. Thus, rates of high school graduation must also be a consideration.

A view of this situation is provided in Table 7 where the numbers of 17-year-olds, the numbers of high school graduates, the numbers of first-time degree-credit enrollments, and associated ratios are given for most of the period of interest. The numbers of persons in the 17-year-old cohort, graduates from high school, and persons enrolling for first-time degree credit all increased substantially during the period 1961 to 1971. However, most of the ratios shown at the right in Table 7 increased as well. That is, using either the number of high school graduates or the number of 17-year-olds in a given year as the base, the ratios of college enrollment increased during the period 1961 to 1971. The source of the increase appears to be an increase in enrollment in two-year colleges, since the ratios for four-year colleges did not increase.

A slightly different view of the college-bound population is given by considering only immediate college entrants (high school graduates who were enrolled in college in October of their year of graduation). Figure 3 shows trends in the percentages of 18-year-olds who, in October of each year, had graduated from high school and/or enrolled in college. Also shown is the percentage of high school graduates entering college. Evident in the data of Figure 3 are fluctuations not indicated in Table 7. The reasons for the differences may be related to "stopouts," i.e., students who delay college entry rather than enrolling in the fall immediately following graduation from high school. There is no reason to believe that in recent years

Figure 2. Numbers of High School Graduates and Immediate College Entrants*



*Adapted from McCandless (1975).

Table 7. College Enrollments During the Period of the SAT Score Decline as a Function of the High School Cohort and Institution Type

Numbers (in thousands)

			First-time degree-credit enrollment (fall)		
School Year	17-year-olds[a]	High school graduates	4-year[c]	2-year[e]	all[c]
1960-61		1,971	774	244	1,018
1961-62	2,768	1,925	770	260	1,030
1962-63		1,950	774	272	1,046
1963-64	3,001	2,290	903	322	1,225
1964-65		2,665	1,041	401	1,442
1965-66	3,515	2,672	989	389	1,378
1966-67		2,679	992	447	1,439
1967-68	3,521	2,702	1,076	554	1,630
1968-69		2,829	1,107	642	1,749
1969-70	3,727	2,896	1,126	654	1,780
1970-71		2,943	1,096	670	1,766
1971-72		3,043			
1972-73					
1973-74					

Ratios

School Year	HS grads./ 17-year-olds	4-yr. Coll./ HS grads.	2-yr. Coll./ HS grads.	All Coll./ HS grads.	4-yr. Coll./ 17-year-olds	2-yr. Coll./ 17-year-olds	All Coll./ 17-year-olds
1960-61	.69[f]	.39	.12	.52			
1961-62	.70	.40	.14	.54	.28	.09	.37
1962-63		.40	.14	.54			
1963-64	.76	.39	.14	.54	.30	.11	.41
1964-65		.39	.15	.54			
1965-66	.76	.37	.15	.52	.28	.11	.39
1966-67		.37	.17	.54			
1967-68	.77	.40	.20	.60	.31	.16	.46
1968-69		.39	.23	.62			
1969-70	.78	.39	.23	.61	.30	.18	.48
1970-71	.76[f]	.37	.23	.60			
1971-72							
1972-73							
1973-74							

Footnotes for Table 7

a. National Center for Educational Statistics (NCES), Digest of Educational Statistics, 1971, Table 68, page 51.
b. NCES, Projections of Educational Statistics, 1972, Table 20, page 47.
c. Ibid., Table 14, page 36.
d. Ibid., Table 15, page 37.
e. Ibid., Table 16, page 38.
f. NCES, Digest of Educational Statistics, 1971, Table 68, page 51 (given as percent of 18-year-olds).

---

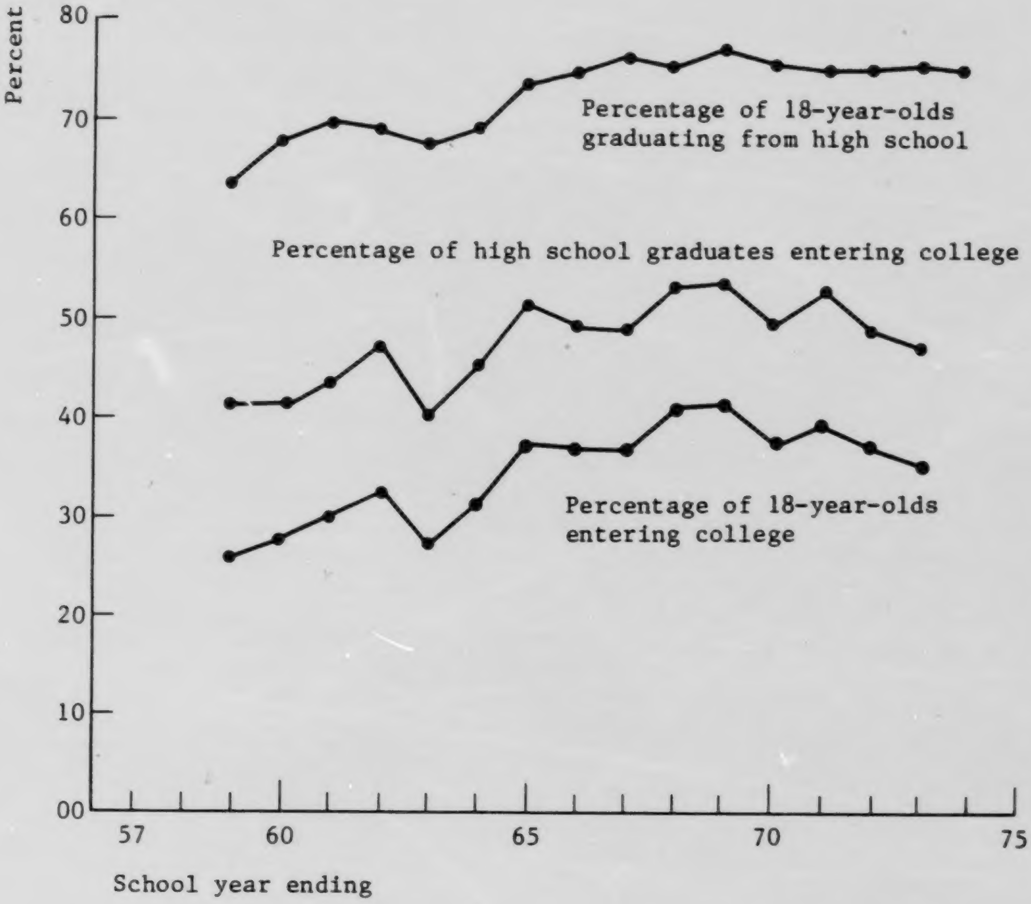
the proportion of stopouts has increased, and this stopout increase would be supported by the downturn in the percentages of high school graduates immediately entering college shown in Figure 3 for the period 1971 to 1973.

Minorities

It is well known that the national college enrollment of minorities has increased substantially during the last 15 years. However, in relation to national college enrollments as a whole, the increase in minority enrollment is such a small proportion that it is unlikely that a significant part of the score decline could be attributed to this increase.

The increase in minority enrollment is documented from a number of scores. Table 8 shows total college enrollment, minority enrollments, percentages of the total represented by minorities, and percentages of high school graduates enrolled in college in terms of total, white, and nonwhite classifications. As suggested in notes to the table, these data should be interpreted with caution. The percentages of nonwhite graduates enrolling in college have wide fluctuations resulting from small census samples. The figures on percent minority enrollment of total college enrollment is corroborated by data collected by the American Council on Education (ACE) for the same period. ACE has developed norms for entering college freshmen annually for a number of years in terms of a wide range of variables, including racial composition. For the freshman class beginning in fall 1966, 10.8 percent in public four-year colleges were either black or American Indian (ACE, 1967, p. 22). In 1972, this figure had increased to 15.9 percent (ACE, 1972, p. 33). That such increases in minority enrollments were not limited to public institutions is also indicated. In 1966, 2.4 percent of students in private universities were black or American Indian. In 1972, this figure was 5.4 percent. Thus, whether one uses the data of Table 8, or that reported by ACE, a proportional increase in minority enrollment of only 4 to 5 percent during the years of the score declines is indicated. Such a small proportional increase in minorities could have only a minimal impact on mean scores. Additionally, it should be noted that for the academic years 1971-72

Figure 3. Rates of High School Graduation and Immediate College Entry for the 18-Year-Old Cohort*



*Adapted from McCandless (1975).

Table 8. College Enrollments of Minorities

Year	Total college enrollment[a] (in thousands)	Total minority enrollment[a] (in thousands)	Minorities as % of total[a]	Percentages of high school graduates enrolled in college: Class	Total	White	Nonwhite
1962				1962	49	51	34
1963				1963	45	46	38
1964	4,643	234	5.0%	1964	48	49	39
1965	5,675	274	4.8	1965	51	52	43
1966	5,999	282	4.7	1966	50	52	32
1967	6,401	370	5.8	1967	52	53	42
1968	6,801	435	6.4	1968	55	57	46
1969	7,435	492	6.6	1969	53	55	37
1970	7,413	522	7.0	1970	52	52	48
1971	8,087	680	8.4	1971	53	54	47
1972	8,313	727	8.7	1972	49	49	48[c]
1973				1973	47	48	35[d]
1974				1974	48	47[e]	51[e]

a. From F. Newman et al., The Second Newman Report: National Policy and High Education, 1973, page 143.
b. From The Monthly Labor Review, September 1974, page 50 (Michelotti, 1974), and The Monthly Labor Review, August 1975, page 35 (Young, 1975). The percentages for the class of 1974 are from the latter. These percentages are based upon civilian, noninstitutionalized high school graduates enrolled in college in the fall term immediately following their graduation from high school.
c. The standard error of this percentage is 4.6.
d. The standard error of this percentage is 4.4.
e. Young (1975) notes: "The proportion of blacks, which had dipped in 1973, returned to the level of the early seventies and was not significantly different from that for whites." (p. 36)

through 1974-75 no substantial change in the proportion of minorities taking the SAT occurred.

Low-SES Students

More important than increases in minority enrollments may be increases in enrollment of low-SES students. Some of the best available evidence on college attendance patterns of low- and middle-SES students during the past several years has been compiled by Cross (1971). Table 9, adapted from Cross, shows data from three different studies spanning the period 1957 to 1967. Although such comparisons should not be interpreted too rigorously, Table 9 would indicate that the proportion of low-SES students attending two-year and four-year colleges has been increasing more rapidly than the proportion of high-SES students. Note that the attendance rates for high-ability, high-SES males remained practically constant at about 90 percent from 1957 to 1967. The same constancy is observable in the second and third ability quarters for the high-SES group. A further comparison of this type will be possible for 1972 when follow-up data from the National Longitudinal Study (NLS)[5] have been analyzed. Since the correlations between SES and test scores is well known, these shifts in college-going patterns by SES groups represent compelling evidence for the part of the decline in average scores occurring between 1963 and 1967. Since about 1971, however, the data of Figure 3 indicate a decrease in proportions of cohorts entering college. Unless there was a concomitant decrease in high- and middle-SES students going on to college, a proportional increase in low-SES students during this later period would not appear possible for immediate college entrants.

Low-Ability Students

Cross (1971) observed that there may have been an increased tendency for low-ability students to consider college, regardless of SES. In describing this trend, Cross defined a new class of students that she called the "New Students." Cross states:

> "The definition of New Students as those scoring low on traditional tests of academic ability may at first seem unacceptable to some readers. Therefore, let me make explicit the reasons for this decision. The primary reason is the very obvious one already discussed. Young people who have not considered college in the past but who are newly entering college in the 1970's are distinguished more by low test scores than by any other single measure available, including race, sex, and socioeconomic status." (p. 14)

THE SCHOOLS

One of the most popular explanations for the decline in SAT and ACT test-score averages is that students graduating from American high

Table 9. Percentages of High School Graduates Attending Two- or Four-Year Colleges[a]

Ability quarter	Socioeconomic quarter 1 (low)			2			3			4 (high)		
	1957[b]	1961[c]	1967[d]	1957	1961	1967	1957	1961	1967	1957	1961	1967
Male:												
1 (low)	6	9	33	12	14	30	18	16	29	39	34	57
2	17	16	43	27	25	39	34	36	55	61	45	61
3	28	32	60	43	38	69	51	48	68	73	72	79
4 (high).	52	58	75	59	74	80	72	79	89	91	90	92
Female:												
1 (low)	4	8	25	9	12	28	16	13	36	33	26	37
2	6	13	28	20	12	36	26	21	50	44	37	67
3	9	25	44	24	30	48	31	40	68	67	65	77
4 (high).	28	34	60	37	51	73	48	71	83	76	85	93

a. Adapted from K. Patricia Cross, Beyond the Open Door, 1971.
b. 1957 graduates, with 1964 follow-up; Sewell and Shah (1967).
c. 1961 graduates, with 1962 follow-up; Schoenfeldt (1968).
d. 1967 graduates, with 1968 follow-up; ETS Growth Study data analysis by Thomas Hilton.

schools are less capable than students in previous years. Two recent and comprehensive research efforts (Armbruster, 1975; Harnischfeger et al., 1975) contain, among other things, useful information concerning the schools. Trends toward promoting and graduating students who have not mastered essentials along with reductions in course requirements are usually the focus of explanations pointing to the schools.

Proportions of Cohorts Graduating

In 1959, slightly more than 60 percent of persons 18 years old in the United States were graduating from high school; in 1973, this percentage was well over 70 percent. Armbruster (p. 84) reports U.S. Census data showing that the percent of cohort kept back one year and/or dropped out by the eleventh grade decreased from 34.9 percent in 1960 to 16.0 percent in 1970. These percentages indicate that students are promoted through the grades who, in previous years, were kept back until they knew the subject matter. Figure 3 shows trends in percentages of persons graduating from high school over the years from 1959 to 1973. Although there has been an increase in proportion of cohort graduating, the nature of the increase is of special interest. Note in Figure 3 that the major portion of the increase in the proportion of cohort graduating from high school occurred between the years 1963 and 1966. By 1967, the trend had leveled off. Table 7 shows that the proportion of cohort graduating leveled off in 1964. Therefore, although the decline in the means of SAT and ACT test scores that occurred between 1963 and 1966 might be associated with an increased proportion of cohort graduating (or reduced school effectiveness resulting from the increased retention rates) during those years, the continued decline in scores beyond 1966 would appear not to be. Even with a decline in the proportion of cohort graduating, a softening of the requirements for graduation could have contributed to the declines in SAT and ACT mean scores. Harnischfeger et al. (1975) have assembled information suggesting that fewer of the basic courses are now required in many schools.

Abilities of High School Graduates

The best data for the analysis of high school graduate ability trends are those obtained from comparable national random samples selected and tested in different years. Unfortunately there have been few of them because studies of this type are expensive. The studies that have been conducted are often questioned on the grounds of sample bias or other methodological problems. Still, the care with which national studies are conducted should make them better indicators than samples of students obtained with no control whatever--such as for the SAT and ACT.

In one study of controlled high school samples, Flanagan and Jung (1971) investigated possible changes in reading comprehension, an area of basic importance for all types of academic development, covering the period from 1960 to 1970. As a part of the well-known Project TALENT study, over 400,000 high school students representing a random sample of 1,353 different schools were administered a 48-item reading comprehension test in 1960. Ten years later the identical reading test was administered to more than 12,000 high school students in 134 schools, this smaller sample having been chosen in the same way as the 1960 sample. Over the 10-year period, June 1960 to 1970, the mean score obtained from this 48-item test increased slightly from 30.81 to 31.25. As might be expected, Flanagan and Jung had difficulty in making arrangements with the schools in 1970 that had been in the original 1960 sample. Some of the schools no longer existed, and biases of several types undoubtedly affected the results. Accordingly, the results may be questioned. But this kind of research evidence represents more control of possible confounding influences than casual observations of SAT mean scores over the years.

In a second study of controlled high school samples, Schrader and Jackson (1975) obtained special samples for the norming of the PSAT/NMSQT. The PSAT/NMSQT is an instrument similar to the SAT but taken, usually, by high school students in their junior year. The scale of the PSAT/NMSQT extends from 20 to 80 rather than from 200 to 800 as in the SAT, but PSAT/NMSQT scores are equivalent to SAT scores if multiplied by a factor of 10. The high school students who normally take the PSAT/NMSQT are, on the average, more capable than the average high school students. Schrader and Jackson, however, used special PSAT/NMSQT norming samples for the years 1960, 1966, and 1974 for their study. In 1974, 17,770 juniors in 142 high schools were selected in approximately the same way as the 1960 and 1966 samples. As in the two previous norms studies, the schools invited to participate in the 1974 norms study were chosen to be representative of all high schools in the United States. There were some differences in the selection process, however. In 1960, more than seven-tenths of the schools invited provided PSAT/NMSQT data for their juniors. In 1966 and 1974, six-tenths of the schools invited provided usable data. Accordingly, the samples did not compare perfectly, and the possibility of differential sampling bias exists. When mean PSAT/NMSQT scores were computed for the verbal and mathematical sections in the three years of the Schrader and Jackson study, the results were as shown in Table 10. These results support neither the Flanagan and Jung finding of no change nor the opposite contention that high school abilities are steadily declining. Table 10 shows that PSAT/NMSQT-verbal means increased between 1960 and 1966 and then decreased in 1974. PSAT/NMSQT-mathematical means show an opposite trend, decreasing from 1960 to 1966 and increasing from 1966 to 1974.[6]

Table 10. Average PSAT/NMSQT Scores for High-School Juniors from Three Norming Samples[a]

Year	PSAT/NMSQT-verbal	PSAT/NMSOT-mathematical
1960.	34.2 (342)	38.7 (387)
1966.	35.8 (358)	37.8 (378)
1974.	34.3 (343)	38.6 (386)

a. Equivalent mean PSAT/NMSQT scores, in terms of the SAT scale, are given in parentheses.

Bracken (1975, p. 42) reports results of national renorming studies of the Iowa Tests of Basic Skills (ITBS) carried out in 1955, 1964, and 1971 (original source: Hieronymus and Lindquist, 1974). These norming studies were conducted for grades 3 through 8; however, by adding four years to the grade 8 data, estimates of the abilities of grade 12 students for the years 1959, 1968, and 1975 may be derived. Some of the more relevant norm changes are given in Table 11. The 1964-71 interval in Table 11 would suggest that grade 12 students at both the 50th and 90th percentiles declined in traditional reading and mathematical skills (but not in vocabulary) between 1968 and 1975.

Bracken also reports (pp. 45-47) on renorming studies of the Comprehensive Tests of Basic Skills (CTBS) and the Iowa Tests of Educational Development (ITED). The CTBS studies were for grades 2 through 10 and covered the period 1968 to 1973. Lower scores were obtained in 1973 for all upper grades in all tests of the CTBS (reading, mathematics, language, and study skills), and the decline was larger the higher the grade. Extrapolation of the grade 10 data suggests a decline in the skills tested by the CTBS for grade 12 students during the period 1970 to 1975. The ITED norming studies showed gains between 1962 and 1975 for high school students in grades 9 through 12.[7]

Another approach to the examination of national trends is to consider results of statewide testing programs. Such data, however, have the disadvantage that migratory influences may affect them. Armbruster (1975) presents data showing convincing declines in reading and quantitative ability for grade 12 students in California between 1969 and 1973 (p. 59), for grade 8 students in Iowa between 1965 and 1972 (p. A-21), for grade 7 students in Michigan between 1970 and 1973 (p. A-23), and for grade 6 students in New York between 1966 and 1972. Nevertheless, Armbruster notes that some states have experienced increases in student performance (p. 49).

Perry and Swanson (1974) studied results from the Minnesota Statewide Testing Program in which almost all eleventh-grade students are tested each year. The test, a version of the Ohio State Psychological Examination, produces a single score measuring primarily verbal comprehension and reasoning. Both the mean and median scores have been declining consistently between the 1968-69 and 1972-73 academic years. Of special interest is that Perry and Swanson also observed a substantial decrease in the number of high scores.

Table 11. Changes in Grade Equivalents for Grade 8 Students on Selected ITBS Tests[a]

Students at	Interval	Vocabulary	Reading	Mathematics
50th percentile.	1956-64	-0.8	2.5	1.6
	1964-71	0.4	-2.1	-3.1
90th percentile.	1956-64	-1.4	-0.5	4.1
	1964-71	0.0	-1.5	-2.1

a. Adapted from Bracken (1975), p. 42.

Despite its questionable generalizability, the comparison of performance, over time, of students within individual schools is of interest--especially when information is also available on important factors that have existed concurrently with the longitudinal comparison. Hawes (1972) reported declines in SAT scores for a stable, high-SES New York state high school over a 10-year period. Between 1962 and 1972, the SAT-verbal median score declined from 575 to 522, even though the town population was unusually stable during this period. Other individual high schools, however, have reported that test scores have increased or have been stable, despite a national decline.

Although much of the available evidence on the performance of high school students during the period of the SAT and ACT score declines would indicate a national decline in traditional academic skills, the lack of consistency in results suggests caution in placing the burden on the schools. More definitive data will come, in time, from the studies of the National Assessment of Educational Progress (NAEP). Baseline data were collected in 1971, and some limited follow-up data have been collected. A 1975 follow-up on writing skills indicates a decline since 1971. No follow-up data are as yet available from the NAEP on reading abilities of the type measured in the SAT. National random samples, like those used by the NAEP, are essential for resolving issues concerning trends in the abilities of the national population.

SOCIETAL FACTORS

Some would argue that changes in the school population are merely reflections of changes in American society. Trends in the adolescent subculture toward drug usage, alcoholism, television watching, and similar societal factors could also have a detrimental effect upon the development of academic skills. It must also be remembered that the children of the period being examined are part of an unusual population upsurge.

There is some research evidence to indicate that there may have been subtle factors operating, especially in the post-World War II boom population. The well-known negative correlation between family size and test performance (children from larger families tend to score lower, on the average, than children from smaller families on almost all kinds of mental tests) is one such factor. Since family size was increasing during the period of the post-World War II population surge, the relationships between family size and test scores would predict a decline in scores for the high school cohort beginning in about 1963, precisely when the SAT decline did, in fact, begin. Zajonc (1975) has shown how mean birth rank, another changing characteristic of the 1946-1956 birth years, correlates with the SAT and other score declines. Grabill, Kiser, and Whelpton (1958) noted that the population increase was disproportionate by SES groups. High- and middle-SES families contributed more to the boom than did low-SES families. But an increase in the higher-SES part of the population in the late 1940s would have tended to produce higher rather than lower scores in the mid-1960s because of the correlation of SES and test performance. And one cannot ignore a possible impact of the GI Bill of Rights. Following World War II, for the first time in American history, large numbers of persons from traditionally nonacademic families went to college. Their children were represented in the SAT and ACT mean score declines. With such an assortment of possible influences operating, it would be surprising if average scores on tests like the SAT and the ACT had remained stable during the period of the declines.

Undoubtedly the rise of television during the same period also had some impact. Just what that impact has been, however, is uncertain. Roger Farr, a recognized expert on the development of reading abilities, has argued that television may *enhance* reading ability. In a comprehensive review of reading abilities by Farr (1975) at several population levels, it was suggested that reading ability in both public and private schools increased from the 1930s until about 1965. A slight decline in reading ability was observed after 1965, but students in 1975 still read better than did students in 1945.

Another view of societal trends in America is offered by the renorming of the Stanford Binet test in 1970 by Thorndike (1975). The Stanford Binet was normed, originally, in 1937, and Thorndike's new norms indicate a gain in the ability of the national population since 1937. Thorndike notes:

"The general rise in IQ level is not surprising. Such a rise has been reported by test-makers on all sorts of ability tests both in this country and abroad." (p. 6)

With respect to Thorndike's contention that abilities have also been increasing abroad, some reading data from England are of interest. Start (1972) reported a study of 11- and 15-year-old English children's reading abilities between 1948 and 1971. Two tests were administered to a stratified random sample of students. The results, presented in Table 12, indicate a gradual increase in reading ability in England between 1948 and 1971. These data on a non-U.S. population are particularly interesting vis-a-vis the hypothesis that time spent watching television is depriving children of valuable reading time and causing a decline in reading skills. Television has also been popular in England.

Table 12. Mean Reading Scores of 11- and 15-Year-Old English Children since 1948[a]

Age	Test	1948	1952	1955	1956	1960	1961	1964	1970	1971
11 years	WV[b]	11.6	12.4	--	13.3	--	--	15.0	14.2	--
	NS6[c]	--	--	28.7	--	29.5	--	--	29.4	--
15 years	WV[b]	20.8	21.2	--	21.7	--	24.0[d]	--	--	23.5
	NS6[c]	--	--	42.2	--	44.6	--	--	--	44.6

a. From Start (1972). Dashes indicate no testing that year.
b. Watts-Vernon Test (35 items).
c. National Survey Form Six (60 items).
d. Mean of two estimates.

CONCLUSIONS

The research evidence that has been presented allows no firm conclusions concerning the declines in mean scores of the SAT or of the ACT. The declines in these two widely used tests, however, appear to be of about the same order of magnitude. Since these two tests are administered to the majority of the college-bound population, and since considerable evidence exists to indicate that the college-bound population has changed over the period of the score declines, it is suggested that those students considering college attendance have indeed changed in terms of traditional academic skills.

The picture for high school students and the national population as a whole is less clear. Considerable evidence suggests a decline in reading and other abilities of high school students, but other data suggest no change at all, and still other research would indicate that an increase in the abilities of high school students occurred during the period of the SAT and ACT mean score declines. One must conclude that the definitive data on this issue are not yet in. The NAEP efforts represent the best hope for resolving this question.

Until it can be determined whether or not there has, in fact, been a change in the abilities of high school graduates, causal analyses seem inappropriate. And because the apparent change in the abilities of the college-bound may possibly also reflect a change in the base population, only speculative observations of the causes for the changes in the college-bound population can be made.

Nevertheless, the cumulative research evidence would indicate that a number of factors, rather than any single cause, have probably contributed to the score decline. The college-bound population, during the years between about 1960 and 1975, appears to have increased in its representation of low-SES families, the mix of SAT user colleges has changed to include larger numbers of two-year colleges and proportionally fewer of the top-ranked four-year colleges and universities, the numbers of students repeating the SAT have declined, and the gains made by those who do repeat have decreased.

FOOTNOTES

1. The December 1973 population of examinees is not necessarily similar to the December 1963 population of examinees because of changes in college and scholarship requirements and because of changes in SAT administration patterns.
2. Carnegie Commission on Higher Education, A Classification of Institutions of Higher Education. New York: McGraw-Hill, 1974.
3. Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualification Test.
4. If there has been a drift in the SAT scale, as discussed previously, then a similar drift would seem to have occurred in the ACT scale also.
5. The NLS is being conducted by the U.S. Office of Education with the assistance of Educational Testing Service and Research Triangle Institute.

6. To the extent that there has been a drift in the SAT scale as discussed on pages 6 and 7, its correction would affect the results obtained by Schrader and Jackson because the PSAT is linked to the SAT scale using the same scale equating methods.
7. In an earlier study by Schrader (1968), comparisons of renorming studies of several national tests also suggested gains, but, these comparisons were based on children up to only grade 8 in the early 1960s.

REFERENCES

American College Testing Program, ACTivity. Iowa City, Iowa: American College Testing Program, October 1975

American Council on Education, National Norms for Entering College Freshman-Fall 1966. ACE Research Reports, Vol. 2, No. 1, 1967

American Council on Education, The American Freshman: National Norms for Fall, 1972. ACE Research Reports, Vol. 7, No. 5, 1972

Angoff, W. A. (Ed.), The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests. New York: College Entrance Examination Board, 1971(a)

Angoff, W. A., "Scales, Norms, and Equivalent Scores," in R. L. Thorndike (Ed.), Educational Measurement, (2nd ed.). Washington, D.C.: American Council on Education, 508-600, 1971(b)

Armbruster, F., The U.S. Primary and Secondary Educational Process (Final Draft). Croton-on-Hudson, New York: The Hudson Institute, July 14, 1975

Bracken, P. J., "A Longitudinal Study of Academic Achievement Levels," in F. Armbruster, The U.S. Primary and Secondary Educational Process (Final Draft). Croton-on-Hudson, New York: The Hudson Institute, July 14, 1975

Carnegie Commission on Higher Education, A Classification of Institutions of Higher Education. New York: McGraw-Hill, 1974

Cleary, T. A., "Test Bias: Prediction of Grades of Negro and White Students in Integrated Colleges." Journal of Educational Measurement, 1968, No.5, 115-124

College Entrance Examination Board, College-Bound Seniors, 1974-75. New York: College Entrance Examination Board, 1975

Cross, K. P., Beyond the Open Door. San Francisco: Jossey-Bass, 1971

Farr, R., Reading, Now and Then (in progress). See New York Times, March 18, 1975, p. 28

Ferguson, R. L., and Maxey, E. J., Trends in the Academic Performance of High School and College Students. Unpublished manuscript. Iowa City, Iowa: The American College Testing Program, 1975

Fincher, C., "Is the SAT Worth Its Salt?" Review of Educational Research, Summer 1974, Vol. 44, No. 3, pp. 293-305

Flanagan, J. C., and Jung, S. M., Progress in Education: A Sample Survey (1960-1970). Palo Alto, Calif.: American Institute for Research, December 1971

Grabill, W. H., Kiser, C. V., and Whelpton, P. K., *The Fertility of American Women*. New York: Wiley, 1958, pp. 384-388

Harnischfeger, A., et al., *Test Score Declines: A Review of Evidence and Possible Grounds for Interpretation* (Draft). Chicago, Ill.: CEMREL, Inc., November 1975

Hawes, G. R., "The Decline of the SATs." *Change*, November 1972

Hieronymus, A. N., and Lindquist, E. F., *Manual for Administrators, Supervisors, and Counselors,* Levels Ed., Forms 5 and 6, *Iowa Tests of Basic Skills*. Boston, Mass.: Houghton Mifflin, 1974

Marco, G. L., and Stern, J., *Investigation of the "Repeater" Hypothesis Concerning SAT Score Declines*. Princeton, N.J.: Educational Testing Service, 1974

Marco, G. L., and Stewart, E. E., *A Test of the Adequacy of Several Score Equating Models When Samples and Test Characteristics Are Systematically Varied. A Proposal for Research and Development*. Princeton, N.J.: Educational Testing Service, January 30, 1975

McCandless, S. A., *The SAT Score Decline and Its Implications for College Admissions*. Paper presented at the 1975 Western Regional Meeting of the College Entrance Examination Board, January 1975

Michelotti, K., "Employment of High School Graduates and Dropouts, October 1973." *Monthly Labor Review*, September 1974, pp. 48-52

Modu, C. C., and Stern, J., *The Stability of the SAT Score Scale* (ETS RB-75). Princeton, N.J.: Educational Testing Service, March 1975

National Center for Educational Statistics, *Projections of Educational Statistics to 1981-82* (1972 ed.). Washington, D.C.: U.S. Government Printing Office, 1973. DHEW Publication No. (OE) 73-11105

National Center for Educational Statistics, *Digest of Educational Statistics* 1971. Washington, D.C.: U.S. Government Printing Office, 1972. DHEW Publication No. (OE) 72-45

Newman, F., et al., *The Second Newman Report: National Policy and Higher Education*. Report of a special task force to the Secretary of Health, Education, and Welfare. Cambridge, Mass.: The MIT Press, 1973

Perry, D., and Swanson, E. O., "Decline in Minnesota College Aptitude Test Scores," in *Decline in standardized test scores: A widespread phenomenon*. Salem, Ore.: Oregon Department of Education, 1974

Schoenfeldt, L. F., Education After High School. *Sociology of Education*, 1968, Vol. 41, No. 4, pp. 350-369

Schrader, W. B., *A Study of the CEEB Candidate Population*. Princeton, N.J.: Educational Testing Service, in progress

Schrader, W. B., and Jackson, R., *PSAT/NMSQT 1974 National Norms for High School Juniors*. Unpublished draft report. Princeton, N.J.: Educational Testing Service, June 2, 1975

Schrader, W. B., *Test Data as Social Indicators*. Princeton, N.J.: Educational Testing Service, SR-68-77, September 1968

Sewell, W. H., and Shah, V. P., "Socioeconomic Status, Intelligence, and the Attainment of Higher Education." *Sociology of Education*, 1967, Vol. 40, No. 1, pp. 1-23

Start, K. B., "Thirty Years of Reading Standards in England." *Educational Researcher*, Vol. 1, No. 11, November 1972

Stewart, E. E., *The Stability of the SAT-Verbal Score Scale* (ETS RB-66-37, CEEB RDR-66-7). Princeton, N.J: Educational Testing Service, July 1966

Temp, G., "Test Bias: Validity of the SAT for Blacks and Whites in Thirteen Integrated Institutions." Journal of Educational Measurement, 1971, No. 8, pp. 245-251

Thorndike, R. L., "Mr. Binet's Test 70 Years Later." Educational Researcher, Vol. 4, No. 5, May 1975

Wilks, S.S., Scaling and Equating College Board Tests. Princeton, N.J.: Educational Testing Service, 1961

Young, A. M., "Students, Graduates, and Dropouts in the Labor Market, October 1974." Monthly Labor Review, August 1975, pp. 33-36

Zajonc, R. B., Intellectual Environment and Intellectual Development: New Data and New Implications. Invited paper presented at the annual meeting of the Midwestern Psychological Association, Chicago, Illinois, May 2, 1975

An Analysis of Textbooks in Relation to Declining SAT Scores

Jeanne S. Chall
Professor of Education
Harvard University
With the assistance of
Sue S. Conard and Susan H. Harris
Research Assistants

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

June 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $5 per copy, as well as a price list for all the appendixes can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

CONTENTS

Acknowledgments 3
Chapter 1. Background of the Study 3
Chapter 2. Findings on SAT Reading Selections 8
Chapter 3. First Grade Readers, Workbooks, and Teachers' Guides. . 13
Chapter 4. Sixth Grade Readers, Workbooks, and Teachers' Guides. . 16
Chapter 5. Sixth Grade Grammar and Composition Textbooks 27
Chapter 6. Sixth Grade History Textbooks 35
Chapter 7. Eleventh Grade Literature Textbooks 43
Chapter 8. Eleventh Grade Grammar and Composition Textbooks . . . 48
Chapter 9. Eleventh Grade History Textbooks 55
Chapter 10. Summary, Conclusions, Recommendations 60
Appendix A 67
Appendix B 73
Bibliography 75

ACKNOWLEDGMENTS

I should like to thank many people who made it possible to do this study.

I wish to thank James Squire, Robert N. Walker, and the editors of Ginn and Company who gave invaluable assistance in selecting the sample of textbooks to be analyzed. I am also indebted to my colleagues Fletcher G. Watson and Arthur Stanley Bolster, Jr. for their counsel on selecting the books to be included in the study.

I am grateful also to the editors of the following publishing houses for their generosity and assistance in making their textbooks available to us:

D. C. Heath & Company
Laidlaw Brothers
The MacMillan Company
Harcourt-Brace-Jovanovich, Inc.
American Book Company
Scott Foresman & Company
Ginn and Company

George Hanford, senior vice president of the College Entrance Examination Board, provided the background information on the Scholastic Aptitude Tests and scores and was always there with a sympathetic ear and a word of encouragement.

I wish to give special warm thanks to Sue Conard and Susan Harris, research assistants and graduate students in human development and reading in the Harvard Graduate School of Education. My thanks, also, to Glenda Bissex for her special assistance in analyzing the grammar and composition and literature textbooks. Their interest, excitement, and meticulous care with the small details as well as the large ideas made our work a joyful experience.

And to Mrs. Ann Cura, administrative assistant in the reading laboratory, Priscilla Hall, Jane Tuchscherer and Willard Osman, my appreciation for their skillfull work on typing the report.

Jeanne S. Chall
Cambridge, Massachusetts
February 14, 1977

BACKGROUND OF THE STUDY

This study was undertaken for the Advisory Panel on Score Decline at the request of Matina S. Horner, president, Radcliffe College, and George H. Hanford, senior vice president, College Entrance Examination Board. It is an exploratory study, undertaken to determine whether any relationship exists between the declining Scholastic Aptitude Test (SAT) scores (1967 to 1975) and the textbooks used by students who took the tests. The limited nature of the study (begun in July 1976) necessarily makes the findings suggestive rather than definitive.

Our assumptions are that textbooks reflect the educational thought and practice of the time in which they are written. In turn, they disseminate theory and practice among teachers and pupils. Thus, most textbooks cannot be too far ahead or too far behind the main educational ideas of their time, particularly if they are to be successful financially. To assure the inclusion of ideas and practices that are current, the large publishing houses select as authors recognized scholars and authorities within the different disciplines. Thus, what is included in textbooks is important for a study of historical changes in educational practices and for a study of curriculum issues.

We further assumed that in studying the relationship between textbooks and SAT scores, representative textbooks from grade 1 to grade 11 should be analyzed.[1] This would permit a series of simulated "longitudinal studies" of the textbook exposure of each of six groups of students taking the SATs--in 1947 and 1955, when their SAT scores were stable; in 1962, a pivotal year;[2] and in 1967, 1972, and 1975 when the SAT scores declined.[3]

More specifically we undertook:

to analyze the difficulty of representative textbooks most likely to have been used by the six cohorts when they were pupils in the various grades of elementary and high school;

to estimate whether any trends in difficulty of the textbooks can be detected and whether these trends may be related to the SAT scores of the six cohorts--the two with stable scores (1947 and 1955), a pivotal year (1962), and the three with declining scores (1967, 1972, and 1975);

to analyze the difficulty of the SAT reading passages taken by the six cohorts, to estimate whether any changes have occurred, and whether these may be associated with changes in the textbooks and in the SAT-verbal scores.

Sampling Decisions

The limited nature of the study necessitated very limited samples for almost every aspect.

Since it was not possible to study each of the stable and declining years, we selected six time periods--1947 and 1955 for the stable SAT years, 1962 for a pivotal year, and 1967, 1972, and 1975 for the declining years. These years were selected by studying the various statistical tables on SAT scores made available to us by George H. Hanford of the College Entrance Examination Board, and in consultation with him.

For each of the selected years only the reading passages and questions from the SAT-verbal test were analyzed. Although one could argue that the vocabulary and analogies give equal or more valuable information to relate to the textbook analyses, we assumed that for a limited analysis that part of the SAT test that most resembles the task required by most textbooks would yield the most useful information.

Since the textbooks in all curriculum areas could not be analyzed, we decided that the following would probably be as representative as any. For grade 1, we selected basal readers, workbooks, and teachers' guides. For grade 6, basal readers, workbooks, and teachers' guides, grammar and composition textbooks, and history textbooks. For grade 11, we selected literature anthologies, grammar and composition textbooks, and history textbooks.

We had planned originally to include science textbooks for grades 6 and 11, but time limitations made it necessary to decide between the history and science texts. The history textbooks were considered to be essential by several curriculum specialists consulted. For a more extensive study it would seem that science textbooks should be included because the SAT reading passages have included many science selections.

Overall, the textbook types selected represented those most likely to be used by the largest number of pupils. They also represented a high concentration of verbal practice and challenge.

The particular grade levels were selected for the following reasons: grade 1 was selected to represent the beginning of formal schooling and the initial acquirement of the reading skill. A great deal is done in teaching the child to read in the first grade and it may well be that a pace is set there. Grade 6 was selected as representing the halfway point of compulsory schooling. It also represents the completion of elementary school. Grade 11 was selected to represent the near completion of high school, and the year in which the cohorts took the SAT.

After discussion with noted textbook publishers, a decision was made to analyze only the best-selling textbooks within each of the selected grades and disciplines. Thus we assumed that if the biggest sellers within the curriculum areas, grades, and copyright years were selected, a high probability existed that our cohorts actually used those particular textbooks. Also, because it is customary for the most successful textbooks to attract followers soon after they reach a peak of popularity, it would seem that they represent others in addition to themselves. Currently, there appears to be a spreading out of the market with more textbooks available and used for each grade and discipline. However, until the early 1960s it was not uncommon for the best selling textbooks to be used by about 50 percent of the children in their intended grades.

The best-selling textbooks within each of the copyright dates were selected in consultation with James R. Squire, senior vice president, and Robert Walker of Ginn and Company, in addition to editors from Ginn's various elementary and secondary school divisions. The selections were also checked with the publishers of the textbooks.[4]

Table 1 contains the distribution of books and publishers used in our analyses. The books are listed by the six cohort years. Thus, for each SAT cohort year (the year the cohort was in the eleventh grade and took the SAT) is listed the year of birth, and the years in which he or she was in grades 1, 6, and 11.

It should be noted that the books are listed as being used three to six years following their copyright dates. This time lag in relation to copyright dates was suggested by the publishers and by several successful authors of elementary and high school textbooks as being the most realistic.

The Analysis

As can be seen from Table 1, 35 textbooks were analyzed, and for more than half of these we also analyzed the correlated teachers' guidebooks and student workbooks.

Table 1. The Textbooks and SATs Analyzed for Six Cohorts

Grade	Year	Reading and literature	History	Grammar and composition
Cohort 1:				
Birth	1931			
1st grade	1937	Publisher G (1930)		
6th grade	1942	Publisher G (1936)	Publisher B (1935)	(Not available)
11th grade	1947	(Not available)	(Not available)	(Not available)
SAT	1947	SAT reading selections		
Cohort 2:				
Birth	1939			
1st grade	1945	Publisher G (1940)		
6th grade	1950	Publisher G (1944)	Publisher B (1942)	(Not available)
11th grade	1955	Publisher C (1949)	(Not available)	Publisher D (1950)
SAT	1955	SAT reading selections		
Cohort 3:				
Birth	1946			
1st grade	1952	Publisher G (See Chall,1967)		
6th grade	1957	Publisher G (1951)	Publisher B (1951)	Publisher C (1952)
11th grade	1962	Publisher C (1952)[1]	Publisher C (1961)[2] Publisher F (1961)	(Not available)
SAT	1962	SAT reading selections		
Cohort 4:				
Birth	1951			
1st grade	1957	Publisher G (1956)		
6th grade	1962	Publisher G (1955-56)	Publisher B (1954)	Publisher C (1955)

(Continued)

Table 1 (Continued)

Grade	Year	Reading and literature	History	Grammar and composition
11th grade	1967	Publisher C (1963)	Publisher F (1964)	Publisher A (1967)
SAT	1967	SAT reading selections		
Cohort 5:				
Birth	1956			
1st grade	1962			
6th grade	1967	Publisher G (1962)	Publisher E (1968)[3]	Publisher C (1964)
11th grade	1972	Publisher G (1968)	Publisher C (1969)	Publisher C (1969)
SAT	1972	SAT reading selections		
Cohort 6:				
Birth	1959			
1st grade	1965	Publisher G (1962)		
6th grade	1970	Publisher G (1965)	Publisher C (1970)	Publisher C (1968) Publisher F (1967)
11th grade	1975	Publisher G (1973)	Publisher F (1970)	Publisher C (1973)
SAT	1975	SAT reading selections		

1. The greater difference here between copyright and probable use dates (10 years as compared to 3 to 6 years generally) occurred because of a delay in shipment of the 1958 edition. Since the readability raw scores of the 1952 and 1958 editions varied little (.3) and their contents were quite similar, we decided to use the 1952 edition for which all the analyses had been completed when the 1958 edition was received.
2. Both books analyzed as it was not clearly determined which was the top-seller.
3. According to the publisher, the 1965 and the 1968 editions were the same. The 1965 was not available, the 1968 was analyzed.

In addition, we analyzed all the reading selections, which ranged from seven to nine passages per test, in the six SATs from 1947 to 1975.

Generally, the analyses of the textbooks and SATs were undertaken to determine the level of difficulty of the text. The workbooks and teachers' guidebooks were analyzed to determine the challenge of the assigned tasks--for example, whether pupils were asked to engage in talking or writing, whether the writing assignments called for simpler or more demanding work.

More specifically, we hoped that the various indices obtained from the analyses would help answer these questions:

1. Was there a change in the difficulty of the SAT reading selections and questions from the stable years (1947 and 1955), a pivotal year (1962), to the declining years (1967, 1972, and 1975)?

2. Was there a change in the difficulty of the textbooks (and demands of the assignments) used by the cohorts of the stable years (1947 and 1955), a pivotal year (1962), and the declining years (1967, 1972, and 1975)?

Some of the more specific questions asked were:

Do various measures of readability, reading maturity, and question complexity indicate a decrease in challenge for textbooks used by the cohorts of 1967, 1972, and 1975--the years of the declining SAT scores? Are the measures higher for textbooks used by the 1947 and 1955 cohorts --those of the rising or stable SATs?

Has there been a change in the amount of reading matter within the same categories of textbooks and grades for the various cohorts and particularly for the stable as compared to the declining years?

Has there been a change in the number and ratio of illustrations? Are there perhaps more illustrations in relation to text in the books used by the declining SAT cohorts?

Has there been a change in the amount and kind of writing assigned in the textbooks, workbooks, and teachers' guides?

Has there been a change in the types of questions asked?

Has there been a change in the amount and kind of instruction given in vocabulary study?

Can changes be detected in the recommendations for additional reading?

CHAPTER 2. FINDINGS ON SAT READING SELECTIONS

The reading passages and questions for the six SAT cohort years (1947, 1955, 1962, 1967, 1972, 1975) were analyzed to determine whether any trends could be found in the challenge they presented. Among the measures used were the Dale-Chall Readability Formula, the Harvard Scales of Readability Assessment, Chall's Reading Stages, and an adaptation of Bloom's Taxonomy for measuring difficulty of questions, etc. (See Appendix B for details on these measures.)

Results

The most striking finding on the SAT reading passages is their extreme difficulty as compared to the textbooks analyzed, particularly those for grade 11. Thus, while most of the SAT passages had a Dale-Chall corrected grade level of 11-12 and 13-15, the eleventh grade textbooks had Dale-Chall corrected grade levels of about 9-10. (See Chapters 7, 8, 9 for analysis of eleventh grade textbooks.)

The other finding that appears to be of interest is the relative stability in the readability scores of the SAT reading passages from 1947 through 1972 (Table 2). However, within this relative stability, some differences can be seen. Thus 1947 and 1955, the two stable years analyzed, are the hardest in readability--each with a corrected grade level of 13-15. From 1962 through 1972 there appears to be a slight decrease of difficulty, with corrected grade levels of 11-12. There appears, however, to be a considerable drop in the Dale-Chall readability scores for the 1975 reading selections. Although the average corrected grade level is still 11-12, it veers closer to a corrected grade level of 9-10.

We also found a trend toward longer passages, with the 1947 and 1955 passages averaging 261 words and the 1975 passages averaging 442 (Table 2). Interestingly, however, the number of questions remained essentially the same-- about five per passage. One wonders whether this has possibly introduced an additional element of ease or perhaps of difficulty.

In our search for measures of question difficulty, we became aware that beginning with the 1967 test, the questions would sometimes ask for the correct answer from a complex set of choices. We present the following, which is not, of course, a direct quote from the 1967 SAT, but tries to capture its style:

All living things need the following to exist:
A) Nitrogen B) Oxygen C) Hydrogen D) Carbon
1. All of these
2. B and C only
3. A, B and C only
4. none of these

Is the intention to measure understanding of what is read or is it to measure logical analysis?

Therefore, we attempted to find a way to measure the qualitative content of the questions and choices. Guidelines for this measure which rate questions and choices on a four point scale--from 1, factual knowledge, to 4, knowledge of universals and abstractions in a field--are found in Appendix B. Each question is judged with regard to the degree of knowledge needed by a student in order to process the questions and choices given for a particular passage. For this scheme we relied heavily on Bloom (1956).

Table 3 presents the findings on the question analysis for the SATs, 1947 to 1975. From it we can note a trend. The passages in the SATs of 1947, 1955, 1962, 1967, and 1972 increase fairly steadily in the degree of knowledge required by the questions, from a mean of 2.5 to 3.1. However, the passages in the 1975 test show a decrease in question difficulty, with an average score of 2.4.

This change in the qualitative difficulty of the questions is seen more vividly in the lower section of Table 3, where the combined per-

Table 2. Relation of questions to Passage Characteristics and Dale-Chall Formula Ratings for SATs

							Dale-Chall Formula			
SAT year	SAT form	Number of reading passages	Average number of words per passage	Number of questions	Average number of questions per passage	Number of running words in passage per question	Sentence length	Dale score (percent un-familiar words)	Raw score	Cor-rected grade level
1947	VSA 3	8	262	45	6	46	31	24	9.0	13-15
1955	DSA 2	9	261	47	5	50	29	23	9.4	13-15
1962	KSA 15	7	362	35	5	72	27	25	8.8	11-12
1967	PSA 13_{cp}	9	461	45	5	92	28	23	8.7	11-12
1972	USA 25	9	418	45	5	84	28	25	8.9	11-12
1975	XSA 1020	7	442	35	5	88	22	22	8.3	11-12

Table 3. SATs. Chall Reading Stages Required by the Reading Passages and Degrees of Knowledge Required by the Questions

SAT Year	1947	1955	1962	1967	1972	1975
Form analyzed	USA 3	DSA 2	KSA 15	PSA 13_{cp}	USA 25	XSA 1020
Average Chall Reading Stage for passages*	4.6	4.1	4.0	4.1	4.8	4.0
Percent Stage 3 for passages	0%	22.2%	14.3%	0%	11.1%	0%
Percent Stage 4 for passages	37.5%	44.4%	71.4%	88.9%	44.4%	100%
Percent Stage 5 for passages	62.5%	33.3%	14.3%	11.1%	44.4%	0%
Average degrees of knowledge required by the questions†	2.5	2.7	2.6	3.0	3.1	2.4
Percent Degree 1 for questions	8.9%	0%	0%	2.2%	0%	11.4%
Percent Degree 2 for questions	48.9%	52.8%	45.7%	22.2%	17.8%	57.1%
Percent Degree 3 for questions	20.0%	30.5%	45.7%	51.1%	53.3%	14.3%
Percent Degree 4 for questions	22.2%	16.7%	8.5%	24.4%	28.9%	17.1%
Combined degrees of knowledge required by the questions:						
Combined percent degrees 1 and 2	57.8%	52.8%	45.7%	24.4%	17.8%	68.5%
Combined percent degrees 3 and 4	42.2%	47.2%	54.2%	75.5%	82.2%	31.4%

*Chall Reading Stages range in increasing levels of maturity from 1 to 5. (See Appendix B for description.)

†Average degrees of knowledge required by the question range in increasing difficulty from 1 to 4. (See Appendix B for description.)

centages of questions of varying degrees of difficulty are presented. As for the average degree of question difficulty, we see an increase from 1947 through 1972 of questions at the two highest degrees--from 42 percent in 1947 to 82 percent in 1972. In 1975, however, there is a drop to 31 percent at the two highest degrees of difficulty, as compared to 82 percent in 1972.

Thus, it appears that according to type of question asked and choices offered, an increase of difficulty could be seen from 1947 through 1972, with a reverse of the trend in 1975. The drop in difficulty in 1975 was found also in the Dale-Chall readability formula measures of the reading passages.

The Dale-Chall and question measures seem to indicate the following: the 1947 and 1955 SATs, those of the stable years, had relatively high Dale-Chall readability scores, but relatively low question difficulty as compared to later tests representing the declining years. The SATs of 1962, 1967, and 1972 tended to have lower readability scores and questions on a higher degree of difficulty. It appears almost as if a balance of difficulty was sought--where the passages were harder, the questions were easier; where the passages were easier, the questions were harder.

This trend seems not to have been continued in the 1975 test, where the difficulty of both the passages and the questions was found to be relatively lower. The Dale-Chall measures on the 1975 passages were the lowest of all the six tests, with a raw score of 8.26 and a corrected grade level of 11-12. The questions also appear to be the easiest of the six tests, with 68 percent of the questions rated 1 and 2 and 32 percent rated 3 and 4 (Table 3).

Another measure was used to classify the passages--Chall's Reading Stages--a measure of the general maturity of reading required of the reader for reading a given selection--from Stage 1, the stage of "saying" the words to Stage 5, the stage of highly sophisticated and selective reading for one's own purposes. (See Appendix B for an explanation of the stages and how they were used to rate the passages.) Briefly, each of the reading passages from the six SATs was rated in terms of whether it required a Reading Stage of 1, 2, 3, 4, or 5. Overall, the average Reading Stage Level for the passages of the six SATs appears to be quite high. Although no consistent trend in the findings from the Reading Stage analysis (Table 3) was found, if 1972 is excluded, there appears to be a decrease in difficulty. The 1947 passages were most demanding, with 63 percent classified as Stage 5, whereas 0 percent of the 1975 passages were classified as Stage 5.

Impressions

After close examination of the SAT reading passages, we were impressed with their difficulty. Indeed, it seems that nowhere in the textbooks being analyzed were the requirements as great. The reading passages in the earlier tests of the stable years--1947 and 1955--seem to require a style of thinking that might be characterized as academic. The passages seemed also to be more abstract.

The later tests--1962, 1967, 1972, 1975--seem to cover topics with which a high school student might be more familiar. Generally, there

Table 4. Subject of SAT Passages by Year

	1947	1955	1962	1967	1972	1975
Biography	1					2
Economics	1	2				
Philosophy	3	1	2	4	1	
Linguistics	1					1
Literature	2		1	1	2	
Music, art.		1		1	1	
Political science		2	1	1	1	
Travel, geography		1			1	
Social science.					1	1
Physical science.		1	2	1	1	1
Biological science.		1	1	1	1	2
Total number of passages for test.	8	9	7	9	9	7

seems to be a trend in the later tests toward making the subject matter more relevant for the student. A comparison of passage topics by year of SAT indicates some change (Table 4). While most of the topics have been and continue to be from the sciences--physical, natural, and social --the early, stable years seem to contain more selections from what are usually considered the more abstract subjects (philosophy, economics, political science) while the later tests had more selections from the less abstract (biography and biological sciences).

There were a few passages from literature in 1947 and 1972, and these were much lower in difficulty--whether judged by Reading Stage Level, readability by scale, or by readability formula. (See Tables I - VI in Appendix A for the various measures on each of the passages from SAT tests in 1947, 1955, 1962, 1967, 1972, and 1975.)

To conclude, findings from the various analyses of the SAT reading passages suggest that all the reading selections from the years studied seem to present a greater challenge than the textbooks analyzed for the eleventh grade.

The analyses also indicated a decrease in challenge of the reading selections from the stable years to the declining years on various measures of readability and maturity of subject matter and treatment. Analyses of questions seemed to reveal the reverse. However, the 1975 SAT reading selections had the lowest challenge on all measures of passage and question difficulty.

CHAPTER 3. FIRST GRADE READERS, WORKBOOKS, AND TEACHERS' GUIDES

The various published studies of first grade reading textbooks delineate a steady decrease of difficulty of beginning readers to the early 1960s (Chall, 1958; Chall, 1967). This lessening of difficulty has been found on such indices as vocabulary load, the number of different or "new" words, the number of repetitions, the percent of high versus low frequency words, and so on.

There were changes also in the teachers' guides. They became more detailed, giving more and more instructions for guiding the pupils' reading. They also gave more instructions for followup activities. However, fewer of these utilized reading and writing. They also put decreasing amounts of emphasis on the teaching of phonics (decoding).

Table 5 presents the relevant data on indices of difficulty and ease for first grade readers published from 1920 to 1962 (Chall, 1967). The first columns present the publication dates, the probable dates used, and the probable years in which pupils who were taught to read by the various editions took the SAT.

From Table 5 we see a consistent decline in the various indices of difficulty and challenge from 1920 to 1962. The vocabulary load, or number of "new" words, decreased steadily, with rather a plateau during the 1950s, but with a new decrease in 1960. The number of new words per 100 running words decreased steadily through 1956, with an increase in 1962. It is interesting to note that while the "reading challenge" kept declining over the 40 year period, the ratios of pictures to text kept increasing.

With regard to the teachers' guides, the analyses indicate a steady increase in the teachers' verbal directions and questions. (See last two columns, Table 5.) This increase in verbal guidance, it should be noted, was taking place during the 1940s, 1950s, and early 1960s, when teachers were advised to divide their classes into two, three, or more instructional groups for reading. Thus, the increasing teacher talk to pupil reading raises the question of whether the teacher was left with enough time to provide each child with the stimulation he or she needed, or whether pupils were given enough reading practice.

From 1920 to 1962 (with a slight reversal in the teachers' guide for 1962), there has been a steady increase in the quantity and specificity of instructions to the teacher. Thus, in 1920 it took 561 words to explain how a typical story was to be taught. There was a steady, steep rise until a peak of 2,300 words of instruction was reached in 1956, with a slight drop to 2,000 in 1962. It should be noted that the length and specificity of the instructions to the teacher paralleled the steady decline in difficulty of the readers in terms of vocabulary and picture loads. It also paralleled the decrease in emphasis on the teaching of phonics (Chall, 1967).

The final column, representing the number of words a child reads per each question asked by the teacher, shows a steady decline, with 47 uninterrupted words read by the child in 1920 to 12 words in 1962.

In summary, for two cohorts in the stable SAT period--1947 and 1955--it appears that their first grade reading programs were substantially more challenging than were those of the declining cohorts.

A more recent analysis of first grade readers published from 1968 to 1975 by Helen Popp (1975) indicates a general reversal of the trends noted above, already seen to a limited extent in the 1962 reader. Particularly evident in the late 1960s and early 1970s were readers with heavier phonics programs (already observed as the beginnings of a trend by Chall in 1967). According to Popp, "since 1967 the changes have been

Table 5. Characteristics of First Grade Readers and Teachers' Guides; 1930-1962 (From Chall, 1967)

Publisher and copyright dates	Years probably used	Years cohorts probably took SAT	Vocabulary load (number of different words in book)	New words per 100 running words	Pictures per 100 running words	Teachers' guide instructional words by teacher per story read	Words child reads per question asked by teacher
Publisher G 1920	1923-26	1933-36	425	2.4	.9	561	47
Publisher G 1930	1933-36	1943-46	282	2.1	1.3	814	26
Publisher G 1940	1943-46	1953-56	178	1.7	1.4	1266	25
Publisher G 1956	1959-62	1969-72	177	1.4	1.6	2300	14
Publisher G 1962	1965-68	1975-78	153	1.9	1.7	2000	12

more dramatic, more obvious....The 1971 edition of the Scott, Foresman Reading Systems retains an emphasis on comprehension and, in fact, indexes 'letter-sound relationship cues' under 'comprehension strategies' ....Level 4 (first grade) includes the recognition and use of letter-sound relationship cues for consonants and vowels." (Popp, 1975, pp. 104-105.) It should be noted that the phonics in the first grade programs of the earlier Scott, Foresman editions included only single consonants. The teaching of vowels in grade 1 indicates significant increase in challenge.

To conclude, the first grade reading programs used by the two stable SAT cohorts, 1947 and 1955, were harder and more challenging. Moreover, if the teacher followed the instructions given in the teachers' guides, these two cohorts probably read more from their readers because the teachers talked less.

The cohorts during the years of declining SAT scores, when in grade 1, were taught from easier readers containing fewer "new" words, a higher ratio of pictures to words, a more limited phonics program, and probably were given less time to read in class because their teachers were spending more time in "verbal guidance."

A change toward greater challenge, started in a rather limited way in the early 1960s, has taken on added momentum since 1967. Although we did not have sufficient time to do a detailed analysis of the latest edition of the first grade Scott, Foresman program, a cursory examination of it, as well as of Popp's classification of reading programs published from 1968 to 1975 (1975), indicates a dramatic increase in challenge--in greater vocabulary difficulty, in greater vocabulary diversity, in the varied content, and in the stronger decoding program. The possible significance of this increase is discussed in the section on summary and conclusions.

CHAPTER 4. SIXTH GRADE READERS, WORKBOOKS, AND TEACHERS' GUIDES

The sixth grade readers analyzed were selected to represent the two "stable" SAT cohorts (1947 and 1955), the pivotal year (1962), and the three "declining" cohorts (1967, 1972, 1975). Table 1 presents the books, copyright dates, the dates most likely in use by the six SAT cohorts, and so on.

The books were analyzed in terms of the degree of challenge they represented. Thus, we analyzed their reading difficulty, the stage of reading development they required, the types of stories, and other aspects of style and content.

The correlated workbooks and teachers' guides were also analyzed for the kinds of practice and methodology suggested to pupils and teachers.

Table 6 presents findings on the readability measures and on the reading stages for the readers most likely used by the six SAT cohorts when they were in the sixth grade. From Table 6 we note little change in the readability scores for the 1936 through the 1962 editions. Although there is some variation in the formula raw score, the Dale-Chall corrected grade levels of the readers averaged fifth to sixth grade. The 1965 edition appears to show an increase in difficulty to a corrected grade level of 7-8.

There appear to be some changes in the readability factors, however, with a steady decrease in the length of sentences from the 1936 to the 1962 editions. The reverse is found for unfamiliar words (Dale score), with a steady increase from 1936 to 1965. Thus, it seems that while the words became harder, the sentences became easier. This helps to explain the relative stability of the formula results, since the Dale-Chall formula is based on these two factors.

Time does not permit us to pursue these changes further. A more comprehensive study would attempt to investigate why these particular changes took place. One possibility might be the growing use of readability measurement as one of the criteria in local and state adoptions of textbooks. (See Chall [1958] for a history of research and application of readability measurement in education.)

One might also ask whether the decrease in length of sentences is part of a general change in our language? (Indeed, adult literature has been found to be decreasing in sentence length since the late 1800s.) Or is it that, in being sensitive to the readability formulas in use by textbook evaluation committees, the sentences were shortened to get a "good score," since sentence structure is easier to change than words.

The reading stages do show some consistent changes. As seen in Table 6, there has been a gradual decline in the reading stage level from 1936 to 1962, but an increase is observed in 1965. This decline from 1936 to 1962 is even more striking when viewed as the percent of stories at various stage levels. Thus, in 1936 the majority of stories (64 percent) were at Stage 3. In the 1944/51 editions, Stage 3 stories dropped to 42 percent and by 1955-6/62 they dropped further to 29 percent. It is important to note the reversal of this trend in the 1965 edition, with 67 percent of stories on Stage 3 level. Thus, there appears to be some evidence for a decrease in challenge during the SAT declining years (with the exception of 1965) as compared with the stable years.

Further evidence of change was sought on the physical and content features--illustrations, pages of reading matter, and types of stories. These findings are presented in Table 7.

As can be seen from Table 7, there is no consistent trend regarding the number of pages of reading matter over time. There is a steady increase, however, in the percent of pictures-from 21 percent of pages containing pictures in 1936 to 38 percent in 1944/51, 47 percent in 1955-6/1962. It should be noted that, again, there is a reversal of the trend in 1965. The extent of pictures dropped to 33 percent.

A definite trend was also found in the amount of literature included in the various editions--a steady decrease with time. In the stable years, about 40 percent of the books contained selections classified as literature (adapted and unabridged). This decreased to 35 percent in 1955-6/1962. Again, as with reading stages and with pictures, the 1965 edition seemed to become more challenging again, with a rise in the literature classification to 86 percent.

The trend is even stronger for the unabridged literature alone, with 35 percent in the 1936 edition, 24 percent in 1944/51, a low of 9 percent in 1955-6/1962 and an increase to 49 percent in 1965.

A similar trend is found for essays and articles: 27 percent in 1936, 6 percent in 1944/51, 2 percent in 1955-6/1962 and an increase in 1965 to 11 percent.

Table 6. Dale-Chall Readability Formula Measures and Chall's Reading Stages of Sixth Grade Readers (Publisher G)

Copyright date	1936	1944/1951*	1955-6/1962*			1965
Book	Basic reader grade 6	Basic reader grade 6	Basic reader grade 6^1		Basic reader grade 6^2	Basic reader grade 6
Grade	6	6	(6^1)	Combined 6^1 and 6^2†	(6^2)	6
Dale-Chall Readability Formula						
Average sentence length	20.1	16.4	(14.0)	13.6	(13.3)	16.5
Average Dale score (Percent unfamiliar words)	6.4	8.9	(8.5)	9.6	(10.7)	9.8
Average raw score	5.6	5.8	(5.7)	5.8	(5.9)	6.0
Average corrected grade level	5-6	5-6	(5-6)	5-6	(5-6)	7-8
Chall's Reading Stage - average	2.7	2.4	(2.1)	2.3	(2.5)	2.8
Percent Stage 2	36%	58%	(91%)	71%	(50%)	25%
Percent Stage 3	64%	42%	(9%)	29%	(50%)	67%
Percent Stage 4	0%	0%	(0%)	0%	(0%)	8%

*The findings for the 1944 and 1951 editions and the 1955-6 and 1962 editions are presented together since there were virtually no changes in them.

†Because the 6^1 and 6^2 readers were analyzed separately, the combined scores are estimated figures for the complete 6th grade program.

Table 7. Physical and Other Content Features Analysis of Sixth Grade Readers Publisher G

Copyright date	1936	1944/1951	1955-6/1962			1965
Book level	Basic reader grade 6	Basic reader grade 6		Basic reader grade 6^1	Basic reader grade 6^2	Basic reader grade 6
Grade level	6	6	Combined 6^1 and 6^2*	(6^1)	(6^2)	6
Total number of reading pages	404	441	574	(288)	(286)	480
Total number of words (approximation)	107,559	79,856	101,096	(50,984)	(50,112)	114,125
Percent of pages with pictures	21%	38%	47%	(46%)	(49%)	33%
Total number of poems	23	0	13	(6)	(7)	17
Percent of total literature	41%	43%	35%	(23%)	(48%)	86%
Percent of adapted literature	6%	19%	26%	(15%)	(38%)	37%
Percent of unabridged literature	35%	24%	9%	(8%)	(10%)	49%
Percent of essays and articles	27%	6%	2%	(0%)	(4%)	11%
Percent of drama	0%	0%	0%	(0%)	(0%)	3%
Percent of stories written specifically for reader	32%	51%	63%	(77%)	(48%)	0%

*Because the 6^1 and 6^2 readers were analyzed separately, the combined scores are estimated figures for the complete 6th grade program.

(Continued)

Table 7 (Continued)

Copyright	1936	1944/1951	1955-6/1962			1965
Book level	Basic reader grade 6	Basic reader grade 6	Basic reader grade 6^1	Basic reader grade 6^2		Basic reader grade 6
Grade level	6	6			combined 6^1 and 6^2*	6
Treatments of stories and articles						
Percent narrative	61%	100%	(96%)	(86%)	91%	65%
Percent expository	39%	0%	(4%)	(14%)	9%	35%
Topics of stories and articles						
Percent home/school/personal experiences	3%	14%	(31%)	(0%)	15%	3%
Percent sports/adventure/fantasy/humor/science fiction	26%	24%	(23%)	(10%)	16%	14%
Percent nature/animals/technology	15%	13%	(7%)	(7%)	7%	5%
Percent nationalism/people of other countries	12%	13%	(0%)	(31%)	16%	26%
Percent history	9%	14%	(27%)	(4%)	16%	6%
Percent biography	15%	14%	(4%)	(24%)	14%	26%
Percent science	3%	0%	(0%)	(0%)	0%	3%
Percent folklore/fairy tales/legends/fables/mythology	17%	8%	(8%)	(24%)	16%	17%

*Because the 6^1 and 6^2 readers were analyzed separately, the combined scores are estimated figures for the complete sixth grade program.

A measure that seems to be the reverse of the amount of literature, particularly unabridged literature, is the percent of stories written specifically for the reader. Although we were not certain that we identified all the stories and articles that fit this category, we present the findings because they appear to be useful. If we use them as an index of "ease," or lack of challenge, then we note a steady increase of ease from 32 percent in 1936 to 51 percent in the 1944/51 editions, and 63 percent in the 1955-6/62 editions. Again, the 1965 edition reverses the trend with 0 percent of stories written specifically for the reader.

Drama is included in only one of the editions--in the 1965. The extent of poetry follows trends found in indices of challenge--highest in the 1936 edition, none in the 1944/51, and somewhat higher in the 1955-6/1962. In 1965, as with other indices of challenge, there is an increase.

The types of stories and the themes also seem to change with the dates of the editions, with more mature treatments and themes included in the earlier, stable years and the less mature in the later, declining SAT score years. Thus, the ratio of expository to narrative stories decreased. In 1936, the ratio was two to three. In 1944/51, all of the stories were narrative, and in the 1955-6/62 editions, 91 percent were narrative. By 1965, the ratio seemed to return to that of 1936, with 35 percent of the selections expository.

With regard to topics, although the trends are not as strong, there does seem to be a change from greater maturity in 1936 to lesser maturity in the 1944/51 and 1955-6/62 editions. As on other measures of challenge, there is an increase in the maturity of topics in the 1965 edition. Thus, stories of sports/adventure, etc., topics of interest to pupils in grade 6, tend to decrease steadily from 26 percent in 1936 to 14 percent in 1965. At the same time, the topics of home/school/personal experiences, ones that are of interest to younger children, increase from 1936 to 1962 and drop again in 1965 (see Table 7).

Analyses of the teachers' guidebooks also seem to reveal a drop in challenge during the years corresponding to the SAT declining years. This can be seen in such varied measures as how the teacher is asked to prepare the children for the reading of their stories, the reasons given to the children for reading, the questions teachers are to ask the children prior to and following the reading of stories, and the amount and kind of writing expected of the pupils after the main reading lesson. Thus, the writing required dropped from 44 percent in the 1936 textbook to 4 percent in 1965 (see Table 8).

Table 9 represents those aspects of the teachers' guides that focus on establishing background for the children and preparation for reading the stories. The trends are not always clear, but here too, there seem to be signs of decreasing challenge over time, at least to 1962; with an increase in 1965.

The same trend can be seen for the "average number of words spoken by the teacher" (as suggested in the guidebook) to prepare the children for reading. The numbers are low in 1936 and high during the 1944/51 and 1955-6/62 years. There is a drop again in 1965. The reasons for reading given to the children show a similar trend--a greater emphasis on what seem to be more mature reasons in 1936, a decrease in such emphasis during 1944/51 and 1955-6/62, with an increase again in 1965.

There is a similar trend for the followup activities suggested in the guidebooks (see Table 8).

Table 8. Followup Activities Suggested in Teachers' Manuals for Sixth Grade Readers (Publisher G)

Copyright date	1936	1944/1951	1955-6/1962			1965
Book	Basic reader grade 6	Basic reader grade 6	Basic reader grade 6^1		Basic reader grade 6^2	Basic reader grade 6
Grade level	6	6	(6^1)	Combined 6^1 and 6^2*	(6^2)	6
Percent of stories with rereading	100%	75%	(64%)	57%	(50%)	92%
Reasons for rereading:						
Percent rereading to find factual information	59%	45%	(43%)	32%	(20%)	27%
Percent rereading for interpretation of tone	0%	22%	(29%)	54%	(80%)	46%
Percent rereading to relate emotionally/personally to story	0%	22%	(14%)	7%	(0%)	9%
Percent rereading to read to others/to dramatize	41%	11%	(14%)	7%	(0%)	18%
Average number of activities per story	2.9	3.6	(4.5)	4.3	(4.1)	6.2
Percent - discussion	16%	42%	(32%)	30%	(27%)	31%
Percent - writing	44%	9%	(6%)	8%	(10%)	4%
Percent - vocabulary study	6%	30%	(32%)	23%	(14%)	54%
Percent of non-print activities	34%	19%	(30%)	39%	(49%)	11%

(Continued)

Table 8 (Continued)

Copyright date	1936	1944/1951	1955-6/1962			1965
Book	Basic reader grade 6	Basic reader grade 6	Basic reader grade 6^1		Basic reader grade 6^2	Basic reader grade 6
Grade level.	6	6	(6^1)	Combined 6^1 and 6^2*	(6^2)	6
Average number of suggested additional readings per story	2.6	1.3	(4.3)	4.3	(4.4)	1.6
Percent of similar stories suggested	66%	38%	(32%)	44%	(57%)	21%
Percent of literature suggested*	17%	8%	(47%)	36%	(23%)	63%
Percent of poems suggested	0%	54%	(15%)	16%	(18%)	5%
Percent of reference works suggested*	17%	0%	(6%)	4%	(2%)	11%

*Because the 6^1 and 6^2 readers were analyzed separately, the combined scores are estimated figures for the complete sixth grade program.

Note: When a general suggestion is given, such as suggesting that a subject be looked up further or that additional similar stores be read, this is counted as one (1).

Table 9. Establishing Background and Preparing Children for Reading Teachers' Manuals for Sixth Grade Readers (Publisher G)

Copyright	1936	1944/1951	1955-6/1962			1965
Book	Basic reader grade 6	Basic reader grade 6	Basic reader grade 6^1		Basic reader grade 6^2	Basic reader grade 6
Grade level	6	6	(6^1)	Combined 6^1 and 6^2*	(6^2)	6
Statements and/or questions spoken by the teacher per story:						
Average number of actual statements and/or questions	1.0	2.8	(5.4)	6.2	(7.1)	3.4
Average number of inferred statements and/or questions	3.5	4.6	(3.4)	4.0	(4.6)	3.6
Average total number of statements and/or questions	4.5	7.3	(8.9)	10.3	(11.7)	7.0
Average total number of words directed by manual for teacher to speak	52	128	(117)	138	(159)	91
Reasons for reading:						
Percent to be entertained by the story	18%	0%	(9%)	9%	(10%)	0%
Percent to relate emotionally to the story	0%	0%	(0%)	5%	(10%)	0%
Percent to find out who, what where, when	50%	38%	(77%)	69%	(60%)	42%
Percent to find out why, how	14%	62%	(14%)	17%	(20%)	50%
Percent to compare, contrast	18%	0%	(0%)	0%	(0%)	8%

*Because the 6^1 and 6^2 readers were analyzed separately, the combined scores are estimated figures for the complete sixth grade program.

The amount of rereading suggested in the teachers' guidebooks is highest in 1936 (rereading is suggested for 100 percent of the stories), drops somewhat in 1944/51, continues to drop in 1955-6/1962 and rises again in 1965.

Although all the findings do not fall directly within this general trend, it seems as if we are dealing with almost a reverse bell-shaped curve, with the greater challenge and difficulty usually found in the 1936 edition, a decline in the 1944/51 and 1955-6/62 editions, and an increase in 1965.

To take just one additional example: Recommendations of reference works are relatively high in 1936 (17 percent), low in the middle years (0 percent in 1944/51 and 4 percent in 1955-6/62) and higher (11 percent) in 1965 (see Table 8).

Pupil activities in the correlated workbooks also show some evidence of decreasing challenge in the 1944 to 1962 editions, with a rise again in 1965 (see Table 10). This may be found in the amount of vocabulary practice--almost none in 1955-6/62, with the highest emphasis in 1965. There also seems to be a high, a low, and then a high again for comprehension skills practiced. As might be expected, the emphasis on phonics at sixth grade level is the reverse of that for comprehension--from relative highs in 1944 to 1962, and a low in 1965. The reversal is logical since phonics is the less mature and challenging activity in the sixth grade.

Among the tasks in the workbooks that have remained relatively stable over time are the amount and kind of writing (Table 10). Generally, the kind of writing required most by the workbooks was underlining, circling, or numbering of answers. In 1944/51 about 59 percent of the workbook exercises were of this kind. This decreased somewhat in 1955-6/62, with 50 percent of the exercises requiring such writing, and a further decrease in 1965 to 45 percent. It appears as if this decrease was balanced by a slight increase in the second most popular way of answering--a single word in a blank space--from 29 percent in 1944/51 to 36 percent in 1965. With regard to the writing of single and multiple sentences, the percentages are low and tend to stay low.

Summary and Conclusions--Sixth Grade Readers

The analyses of the sixth grade readers, teachers' manuals, and pupil workbooks revealed a trend of decreasing challenge from 1944 to 1962. The trend was not as clearcut as for the first grade readers, but there were sufficient signs that the sixth grade readers published in 1936 were also more challenging than those published during 1944/51 and 1955-6/62. Similar to the first grade readers, there appears to be a current trend toward increasing difficulty and challenge as seen in the 1965 edition.

Table 10. Analysis of Workbooks for Sixth Grade Readers (Publisher G)

Copyright	1936	1944/1951	1955-6/1962			1962
Book	Basic reader grade 6	Basic reader grade 6		Basic reader grade 6^1	Basic reader grade 6^2	Basic reader grade 6
Grade level	6	6	Combined 6^1 and 6^2*	(6^1)	(6^2)	6
Total number of pages		96	75	(75)	(75)	127
Estimated amount of expected writing:						
Estimated percent of underline, circle, number the correct answer		59%	50%	(52%)	(49%)	45%
Estimated percent of one word answers		29%	31%	(28%)	(33%)	36%
Estimated percent of one sentence answers		12%	17%	(18%)	(15%)	15%
Estimated percent of more than one sentence answers		0%	2%	(2%)	(3%)	4%
Skills Practiced for sample stories:						
Percent of comprehension		20%	11%	(22%)	(0%)	44%
Percent of phonics, rootwords, and affixes		12%	16%	(13%)	(19%)	3%
Percent of structural analysis of sentence		3%	10%	(0%)	(19%)	0%
Percent of use of context clues		17%	19%	(26%)	(13%)	0%
Percent of vocabulary meaning		20%	3%	(0%)	(7%)	31%
Percent of locating and organizing information and reading for specific purposes		28%	41%	(39%)	(42%)	22%

*Because the 6^1 and 6^2 readers were analyzed separately, the combined scores are estimated figures for the complete sixth grade program.

This trend was found for the levels of reading stages, the proportionate number of pictures to print, the amount and kind of literature included in the reader--particularly unabridged literature, stories written expressly for the reader, and the ratio of expository to narrative writing. The analyses of the teachers' guidebooks and pupil workbooks also revealed a similar trend of more challenging activities in the earlier editions, a decrease in the middle editions, and often an increase in challenge in the 1965 edition.

Several indices showed little change with time--the Dale-Chall Readability Scores, for example were essentially the same in the 1936 to the 1962 editions, although the 1965 showed an increase. Some trends were also noted with regard to a possible increase in unfamiliar words and a decrease in sentence length after 1936. Another index of challenge that did not follow the usual trend was the extent of writing suggested. All the editions seemed to suggest few writing activities, but these also represented a steady decrease from 1936 to 1965. The most common writing activity was to underline, circle, number, or fill in a blank with one word. The writing of more than one sentence was rarely suggested.

CHAPTER 5. SIXTH GRADE GRAMMAR AND COMPOSITION TEXTBOOKS

General Scheme for Analysis

Five sixth grade grammar and composition textbooks were analyzed (copyright dates 1952 to 1968). Unfortunately, these covered only the SAT cohort years of 1962 to 1975. The sixth grade grammar textbooks likely to have been used by two of the stable SAT cohorts--1947 and 1955--were not available for analysis.

A late 1960 version of a best-selling textbook of the stable years was made available and was included in our analysis. However, although the newer edition retained the same author as the requested earlier edition, the content differed. It was written from the viewpoint of the new transformational grammar of Noam Chomsky. This revolutionary view of language and grammar was just published in 1958, in Chomsky's Syntactic Structures. Since Chomsky's new view of grammar was, in all probability, not available to textbook authors before 1958, the late 1960 version of the sixth grade grammar textbook could not be the same as the requested, pre-1958 edition.

The major purpose of the analyses --a comparison of the degree of challenge presented over time, with regard to differences between the textbooks used during the stable years and the declining ones--could not, therefore, be achieved.

Since our analysis did not cover the full range of cohort years, the findings are presented as essentially descriptive and suggestive. Trends observed will be only for the years 1962 to 1975.

Analysis of the grammar and composition textbooks presented still another problem. These books are not "read" as are the readers and history textbooks. Their main purpose is to teach how to write through appropriate knowledge, demonstration, and practice, and to impart knowledge of and skill with the language.

To estimate the degree of challenge they present, overall measures of readability and reading stage development would not be appropriate. Instead, we analyzed only selected portions--examples and illustrations of good student writing, and examples of general good writing from literature, historical speeches, or similar sources.

Both the student and literary examples of good writing were tested for readability by the Dale-Chall Formula and for Chall's Reading Stage Level. In addition, the literary examples were rated on a three-point scale of language complexity (from familiar and informal to formal and difficult) and a three-point scale of content difficulty (from familiar life experiences to highly detailed and technical). (See Appendix B for further details on these rating scales.)

Analyses were made of other features of the books: the total number of reading pages, approximate number of words in the text, and so on. To estimate the importance placed on vocabulary study, and on the study of sentences, paragraphs, and themes, we estimated the percent of the textbook devoted to each of these. In addition, a breakdown was made of the relative emphasis placed on text, exercises, and examples.

Since one of the major purposes of such texts is to improve the students' writing, an estimate was made of the amount of writing required or suggested. Thus, the exercises were classified primarily as verbal, or as requiring some written response--for example, underlining the correct answer, punctuating or capitalizing; inserting a word in a blank; writing a one-sentence answer; writing more than one sentence (including an outline); writing a paragraph, story, report, letter, theme, or term paper.

Our hypotheses were that the readability levels, the Reading Stage Levels, and the content and linguistic difficulty measures would differ for the student and literary examples; that the student examples would tend to be easier--with lower readability and stage level scores, and that the readability and stage levels would rise and fall in relation to the increases and declines of the SAT cohort scores.

A perusal of Table 11 suggests that the pupils who used the more recently published grammar and composition textbooks were exposed to a heavier load of work--the number of reading pages increased from 1952 to 1968, as did the total number of words, the number of pages devoted to vocabulary study, and so on. An increase was also found in the number of pages devoted to exercises--from about one-fourth in 1952 to more than one-third in 1955 and over one-half in 1964 and 1968.

The linguistic complexities seem to remain about the same over the four sample years. Thus the Dale-Chall readability levels for the examples of good student writing remain about the same--at a 5-6 corrected grade level (Table 12). It is interesting to note, however, that similar to the Dale-Chall readability findings for the sixth grade readers, there is a rise in the percent of unfamiliar words and a decline in average sentence length, while the raw scores and corrected grade levels remain the same.

Table 11. Physical and Other Content Features Sixth Grade Grammar and Composition Textbooks

Date	1952	1955	1964	1968	1967*
Publisher	Publisher C	Publisher C	Publisher C	Publisher C	Publisher F
SAT cohort	1962	1967	1972	1975	1975
Total reading pages	319	320	354	354	365
Total words approximation	53,018	64.170	80,648	82,713	96,524
Number of words per reading page	166	200	228	234	264
†Total number of picture pages	96	94	87	77	95
†Percent of picture pages	30%	29%	25%	22%	26%
†Total number of vocabulary pages	104	101	157	142	139
†Percent of vocabulary pages	33%	32%	44%	40%	38%
†Total number of exercise pages	79	120	192	183	174
†Percent of exercise pages	23%	38%	54%	52%	48%
†Total number of sentence, paragraph, and theme pages	38	37	74	85	100
†Percent of sentence, paragraph, and theme pages	12%	12%	21%	24%	27%

*Because the 1942 and 1957 editions requested were not available, the publisher sent this 1967 edition by the same author. It is based on the recently developed theories of transformational grammar.

†Not mutually exclusive.

Table 12. Readability Measures and Chall's Reading Stages
and Linguistic and Cognitive Levels for Sixth Grade Grammar and Composition Textbooks

Date	1952	1955	1964	1968	1967*
Publisher	Publisher C	Publisher C	Publisher C	Publisher C	Publisher F
Student Writing Examples†					
Dale-Chall Readability Formula					
Average sentence length	15.0	13.0	13.5	13.5	11.5
Average Dale Score (percent of unfamiliar words)	5.5	6.8	8.2	8.2	6.5
Average raw score	5.2	5.4	5.6	5.6	5.2
Average corrected grade level	5-6	5-6	5-6	5-6	5-6
Chall's Reading Stage average	2.2	2.5	3.0	3.0	2.2
Percent of Stage 2	75%	75%	0%	0%	75%
Percent of Stage 3	25%	25%	100%	100%	25%
Literary Writing Examples†					
Dale-Chall Readability Formula					
Average sentence length	38.0	27.2	23.0	28.2	12.8
Average Dale score (percent of unfamiliar words)	9.8	7.8	8.7	7.8	7.8
Average raw score	7.0	6.2	6.1	6.3	5.5
Average corrected grade level	9-10	7-8	7-8	7-8	5-6
Chall's Reading Stage average	2.2	2.2	2.0	2.0	2.8
Percent of Stage 2	75%	75%	100%	100%	25%
Percent of Stage 3	25%	25%	0%	0%	75%

(Continued)

Table 12 (Continued)

Date	1952	1955	1964	1968	1967*
Publisher	Publisher C	Publisher C	Publisher C	Publisher C	Publisher F
Linguistic level average.	2.0	2.0	2.0	2.0	2.0
Cognitive level average	1.0	1.2	1.0	1.2	1.8
Exercise sentences					
Dale-Chall Readability Formula					
Average sentence length. . . .	7.5	7.8	8.2	9.8	7.2
Average Dale score (percent unfamiliar words).	8.0	6.2	9.5	7.8	1.0
Average raw score.	5.3	5.0	5.5	5.3	4.2
Average corrected grade level	5-6	5-6	5-6	5-6	4 or below

*Because the 1942 and 1957 editions of the requested books were not available, the publisher sent this 1967 edition by the same author.
†Note: The same examples of student writing are used in both the 1964 and 1968 editions.
**Note: Only three examples could be found for sample of literary "good writing" for 1964.

The examples of literary writing seem to get easier with later publication dates. There is a decrease in readability from a Dale-Chall grade level of 9-10 in 1952 to a 7-8 in 1968. There is a decrease in average sentence length, from 38 words in 1952 to 28 in 1968. As noted earlier, the 1967 book represented a new trend in grammar textbooks (the use of transformational grammar) and was therefore not included in our study of trends.

The readability levels of the exercise sentences are rather consistent (at 5-6 grade level) although there appears to be a drop in 1967 to fourth grade or below.

To summarize the readability findings: it appears that while the amount of reading in the textbooks may have increased, the complexity level of the language decreased. Thus, over time, the pupils may have been working longer on less challenging material.

Other Estimates of Maturity and Complexity

The Reading Stage levels for the examples of student writing (Table 12) seem to show a trend toward higher maturity from 1952 to 1968, if we exclude the 1967 volume.

The Reading Stage levels for the examples from literary writing (Table 12) seem to show no consistent trend over time. Contrary to findings of the readability measures, they are lower than are those for the examples of student writing.

It is difficult at this point to interpret these findings, except that the examples used by the authors tended to stay within the lower Reading Stages, whether they were from student writing or from literary or historical works. This seems to be borne out by the findings on the linquistic and cognitive levels of the literary examples. According to the three point scales devised and used on the literary examples, the cognitive level (1.2) seems to be below the linguistic level (2.0). Neither seems to show a general trend over time.

The remaining levels of challenge studied are concerned with the amount and kind of work required of the student when doing exercises in the textbook (Table 13) and in the correlated workbook (Table 14). Overall, there appears to be very little challenge in the writing tasks assigned or suggested. The great majority, approximately 60 to 70 percent, require only underlining the correct answer or filling in one word. There seems to be little change on these from 1952 to 1968.[5] Writing of paragraphs, themes, stories, reports, and letters seems to decrease over time. These writing requirements are generally low, from about 13 percent of exercises in 1952, and decrease further to 6 percent in 1968.

The status of writing is similar in the workbooks (Table 14). Most of these exercises ask the student to underline, circle, and insert a word or a sentence in the proper spaces. Nineteen percent of the exercises in 1955 called for the writing of more than one sentence. This drops to 7 percent in 1965 and 1968.

Summary

The analyses of sixth grade grammar and composition textbooks covering the years 1962 to 1975 revealed the following.

Table 13. Degree of Writing Required for Exercises in Sixth Grade Grammar and Composition Textbooks

Date	1952	1955	1964	1968	1967*
Publisher	Publisher C	Publisher C	Publisher C	Publisher C	Publisher F
Types of exercises:	N/105	N/134	N/205	N/198	N/178
Verbal only	7%	11%	9%	14%	9%
Exercises requiring writing:					
Underline, choice, punctuate, capitalize	43%	30%	44%	38%	21%
One word	25%	31%	28%	29%	37%
One sentence	9%	15%	7%	11%	11%
More than one sentence	3%	3%	4%	2%	4%
Paragraph	2%	4%	2%	2%	8%
Theme, story, report, letter	11%	6%	6%	4%	10%
Combined:					
Underline and one word	68%	61%	72%	67%	58%
More than one sentence and paragraph and theme	16%	13%	12%	8%	22%

*Because the 1942 and 1957 editions of the requested grammar books were not available, the publisher sent this 1967 edition by the same author.

Table 14. Analysis of Workbooks for Sixth Grade Grammar and Composition

Date	1955	1965*	1968*
Publisher	Publisher C	Publisher C	Publisher C
Total number of pages	158	158	158
Total number of exercises requiring writing	149	115	115
Percent of underline or circle	30%	33%	33%
Percent of one word	30%	37%	37%
Percent of one sentence	21%	23%	23%
Percent of more than one sentence	19%	7%	7%

*Note: These two workbooks are very similar and use many of the same exercises.

The volume of reading matter seems to have increased over the years.

The Dale-Chall readability levels of good student writing examples seem to have remained at essentially the same level--5-6 grade.

The readability levels of exercise sentences tend to be similar to those for good student writing--these are rather consistent at 5-6 grade level, although there appears to be a drop to grade level 4 for the 1967 book.

The readability levels for the examples from literature and historical works decrease over time, with a 9-10 grade level for the textbook copyrighted in 1952, a 7-8 grade level for those copyrighted in 1955 and 1964, and a 5-6 grade level for the 1967 book.

The findings for Reading Stages are not as consistent as for the readability analyses. Generally, there appears to be some increase over time of the Reading Stages for good student writing. The Reading Stages for the literature examples do not show a trend, however.

The writing tasks show a slight trend, with a lessening (with one exception) of more substantial writing tasks such as paragraphs, themes, stories, and reports from the 1952 book to the 1968 one. There is also an increase in such tasks as underlining, circling, numbering, and inserting a word in the provided space. What seems to be of considerable importance is that most of the writing tasks were of the latter type.

To conclude, our analyses of the sixth grade grammar and composition textbooks do not provide firm evidence of decreasing challenge, possibly since we could not analyze the two oldest books used by the stable SAT cohorts. At the same time, our analyses failed to reveal firm evidence of increasing challenge. Indeed, we found a few indices that the textbooks over the declining years presented a reduced challenge to sixth graders. These indices were the decreasing readability levels of the samples of literary writing in the grammar and composition texts; the rather low Reading Stage Levels of these selections, lower in some books than the Reading Stage Levels of samples of good student writing; and the lessening over time of assignments for writing stories, themes, and reports.

In one of the grammar and composition textbooks published in the 1960s, the tone and style of the books to which many of the declining SAT cohorts were exposed is illustrated by the section on examples of good writing, where the authors note that while the selections should be read and appreciated, they should not be considered examples of the type of writing that students should expect to attain themselves. Thus, it appeared as if students were informed, in subtle ways, that they need not strive too hard.

CHAPTER 6. SIXTH GRADE HISTORY TEXTBOOKS

The sixth grade history textbooks that were analyzed cover a 35 year period--from 1935 to 1970. As can be seen from Table 1, the full range of all six cohort years is included.

The analyses undertaken for these books were similar to the ones for the sixth grade readers. In addition, we included analyses of the questions asked in the texts, manuals, and workbooks. Generally, we sought to find changes in the challenges students were exposed to over the various cohort years--and to determine whether any relationship exists between these challenges and the SAT scores.

History and other content area textbooks contribute in at least two ways to the abilities measured by a test such as the SAT verbal test. They provide the basic knowledge, information, and tasks, within a particular area of the curriculum. But they also provide the practice in literacy skills necessary for learning in that particular curriculum area. Thus, while for those who have completed their formal education the major objective for reading history (or science) is to know and learn history and science, for children the history and science textbooks perform still another function. They provide the reading practice necessary for developing further the pupil's cognitive, linguistic, and literacy skills. What should the content and difficulty level of history textbooks be if these two purposes are to be fulfilled?

As far as we know, there are no empirical studies on the optimal levels of difficulty for particular kinds of content materials for children of given ages and abilities. Among the measures used by schools and publishers since the 1940s are readability formulas, the same as those used for reading textbooks. We have, therefore, used a readability formula for one of the measures of challenge.

Other measures of challenge used were physical features of the text--the quantity of print as compared to the number of illustrations--Chall's Reading Stage Levels and analysis of the questions for cognitive difficulty.

Findings

The most striking change found was in the use of pictures and illustrations. As can be seen in Table 15, the percent of pages containing pictures (of total reading pages) rose from 46 percent in 1935 to 87 percent in 1970. There was also an increase in the percent of charts, maps, and graphs, but it was relatively small--from an average of about 4 percent for the stable SAT years to about 12 percent for the declining years.

There are changes in the length of the textbooks--in total number of pages and in approximate total number of words--but the trend is not consistent for all six cohorts. There does seem to be, however, some indication that the earlier books were shorter. Then during the "middle cohorts years" there seemed to be an increase in number of words and pages, with a slight decrease in pages and a marked decrease in words in the declining years.

What these changes can be attributed to is hard to say. They may well be explained by the difference in authors. Or they may indeed represent a trend from a more neatly packaged textbook that pupils are expected to read to another conception of a textbook--one that may be a combination of an encyclopedia and a picture magazine.

The analysis of physical features indicates the possibility of some loss in challenge through an increase of pictures. From 1935 to 1970 there has been an increase of pictures to text from 46 percent to 87 percent. What this increase in pictures may indicate about the general style of the history textbooks will be suggested later in an impressionistic report on these same books.

Table 15. Physical and Other Content Features Sixth Grade History Textbooks

Date	1935	1942	1951	1954	1968(65)*	1970
Publisher	Publisher B	Publisher B	Publisher B	Publisher B	Publisher E	Publisher C
Total number of reading pages. . . .	292	368	419	423	378	369
Total number of words (approximation). . .	60,182.2	77,234.3	154,974.3	156,312	88,059	71,841
Total number of picture pages. . . .	134	191	326	326	278	320
Percent of picture pages to total reading pages. . . .	46%	51%	78%	77%	74%	87%
Total number of charts, maps, graphs	8	20	60	60	32	45
Percent of charts, maps, graphs to total reading pages .	2.7%	5.4%	14.3%	14.3%	8.4%	12.2%

*The 1968 edition was received instead of the 1965 edition with the publisher's statement that they were the same.

The readability measures do not show a strong trend (Table 16). The scores for the six cohorts tend to be quite similar, with four of the corrected grade levels at 5-6, and the last two cohorts at 7-8 grade level. Thus, it appears from the readability scores that there is little difference in challenge when the declining years are compared to the stable ones. There may, in fact, be an increase in difficulty for the last two cohorts.

The Reading Stage Levels are also consistent throughout the years, although there may be a slight decline for the 1951 and 1954 books.

The question analyses did not produce a recognizable trend from the stable to the declining years (Table 17). But there seems to be some increase in the cognitive demands of the questions beginning with the 1965 text.

Impressionistic View of the Sixth Grade History Textbooks

The following, impressionistic, and admittedly subjective, view of the history textbooks is presented to help explain some of the changes we felt were present but which were not fully revealed by the more objective analyses reported earlier.

Generally, our impressions were that these books had changed considerably over the 35 year period from 1935 to 1970--from relatively small books, straightforward and factual, to books that resemble encyclopedias on the outside and picture magazines on the inside, with illustrations on three out of four pages and a matching, somewhat chatty style. The earliest book, copyrighted 1935, has one small illustration of children appearing at the beginning of each unit. All other illustrations are photographs, reprints of paintings, sculptures, museum displays, and so on. These illustrations and the book itself do not seem to be directed to the child. The style of writing is factual and, as with the illustrations, there seems to be no attempt to be "child-related."

Generally, the 1935 edition seems to be a book meant to be read--small in size and designed to impart information. It does not try to entertain. The style of writing assumes that the child will reach for the material rather than having the material accommodate the child. The book gives the impression of having something important to say. This impression is reinforced by the illustrations and by the questions on the information.

The 1942 book has much the same content as that of 1935. One difference is that this edition begins with a very brief story about a little boy being lost. This seems to be the only reference to the fact that this is a textbook for children. The other material is written in a factual way.

Since a photographic copy was used for the analysis, we are not certain at this time whether most or all of the illustrations were black and white. The book includes many drawings that seem closely related to the text. There are also instructions given to the reader on how to look at the pictures. For "A Desert Well," the reader is asked to notice the dog's load. In a picture of drawing water from a well by a windlass, the reader is asked, "Do you see where the rope twists around the pole?"

The 1942 textbook is still small, although the printing is larger than in the 1935 edition. Directions to the child are given at the chapter's end--topics to talk about, things to do--a trip to the library or museum or copying in manuscript a favorite poem, for example.

In summary, although the purpose is not stated, the 1942 edition also appears to be concerned primarily with imparting information to the child. Some concern is shown for child-centeredness by including the first story about the child being lost. But it quickly relates this experience to early man's need for learning skills of survival.

The 1951 textbook is quite different when compared to the two earlier editions. It is a larger book. The print is larger. There are differences in the style of writing and in the illustrations. Many more illustrations are included, some drawn expressly for the book, and they are quite childlike. There are many full pages of pictures, and even songs.

The book is written to the child. Each unit starts with the story of a child who might have lived in that particular time or location. This might be a story book, a book written to entertain. The information is still included, but it is not as easily found as in the earlier textbooks.

The 1954 textbook is almost identical with the 1951 edition, with some updating of the more recent years of history. The 1968 textbook[6] includes much information for teaching, with extensive directions to the teacher. New vocabulary is listed; films and filmstrips to use are included; teaching procedures are delineated.

The style of the book follows the pattern of the 1951 and 1954 editions. It is an attractive book, large and beautifully bound, with large print and much white space. The illustrations are colorful and include cartoons. For the most part, the pictures were drawn especially for the book and these are very child-centered. The book includes "Studies in Depth" aimed at the more able students, and writing assignments are indicated there.

Generally, the 1968 textbook was written for and to the child. In fact, it seems to do almost everything for the child. Does this perhaps leave little for the child to seek for himself? It is informative, but the information may be lost in the narrative style.

The Teachers' Guide to the 1970 textbook gives a long discourse (42 pages) on the discipline of history and on a definition of social studies, including the values and concepts which guide men's search for understanding. Edwin Fenton, Teaching the New Social Studies: An Inductive Approach, is quoted on teaching values. Concept-seeking strategies are emphasized, involving quotes or ideas from Bloom, Bruner, Goodnow, Austin, Estavan, and Piaget. It is interesting to note, however, that in these instructions to the teacher it is pointed out that although the program lends itself to all levels of written expression, it is not necessary for the student to write in order to succeed.

The 1970 edition is a beautiful book, filled with colorful illustrations. The pictures are not as child-centered as in the textbooks previously discussed. Much has been included, and sometimes the organization is somewhat difficult to follow. It seems not to be a book to be read but to be referred to. It seems more like a digest; pages are broken up by inserting points of emphasis, or points for further study. It seems to try to involve the child in its discourse. The writing does not talk down to the child. The style, in fact, seems to reflect that of the more sophisticated young person of our time.

This is not a story book. It is filled with information. Yet it may have so much in it that readers have difficulty in finding what is important. It seems that the book tries to have something for everybody. And in doing so, it may perhaps miss everybody. Generally, from this text, the child may learn much; specifically, he may learn very little.

Table 16. Dale-Chall Formula Readability Measures and Chall's Reading Stages for Sample Passages from Sixth Grade History Textbooks

Date	1935	1942	1951	1954	1968(65)*	1970
Publisher	Publisher B	Publisher B	Publisher B	Publisher B	Publisher E	Publisher C
Dale-Chall Formula:						
Average sentence length.	15.8	13.0	15.6	15.6	12.7	13.7
Dale score (percent unfamiliar words).	7.4	7.1	8.7	8.7	11.7	12.5
Raw score.	5.6	5.4	5.8	5.8	6.1	6.3
Corrected grade level	5-6	5-6	5-6	5-6	7-8	7-8
Chall's Reading Stages†						
Average stage for passage	2.8	2.8	2.6	2.6	2.8	2.8
Percent of Stage 2 .	11%	21%	29%	29%	8%	10%
Percent of Stage 3 .	89%	75%	71%	71%	92%	90%
Percent of Stage 4 .	0%	4%	0%	0%	0%	0%

*The 1968 edition was received instead of the 1965 edition with the publisher's statement that they were the same.

†Chall's Reading Stages range from 1 to 5 in increasing levels of maturity. (See Appendix B for description.)

Table 17. Analysis of Questions for the Sixth Grade History Textbooks

Date	1935	1942	1951	1954	1968(65)*	1970
Publisher	Publisher B	Publisher B	Publisher B	Publisher B	Publisher E	Publisher C
Average number of questions asked per chapter.	13.4	8.5	12.8	12.8	14.6	10.5
Average level of questions asked† . .	1.24	1.56	1.35	1.35	1.58	1.53
Percent of level 1 questions.	76%	44%	65%	65%	52%	57%
Percent of level 2 questions.	24%	56%	35%	35%	38%	33%
Percent of level 3 questions.	0%	0%	0%	0%	10%	10%
Percent of level 4 questions.	0%	0%	0%	0%	0%	0%

*The 1968 edition was received instead of the 1965 edition with the publisher's statement that they were the same.

†Questions are differentiated as to 4 levels of increasing difficulty as follows: Level 1, questions asking who, what, when, where and/or list; Level 2, questions asking why, how, explain and/or describe; Level 3, questions asking to compare and/or contrast; Level 4, questions asking to justify and/or support. (See Appendix B for a fuller discussion.)

We heard recently of the response of a classroom teacher to a parent's inquiry as to why her child does not bring home her history textbook for reading. "Oh," said the teacher, "we don't read our history textbooks, we use them for reference."

Comparable Excerpts from Sixth Grade History Textbooks

As further impressionistic evidence of the changes in the style and tone of the sixth grade history textbooks we present below short excerpts on the same topic--the early training of boys in ancient Sparta--from several of the books analyzed. The first two excerpts are ones that might have been used by the two stable cohorts--those taking SATs in 1947 and 1955. The others might have been used by the pivotal cohort of 1962 and by the declining cohorts of 1972 and 1975. It seems to us that these short selections further illustrate some of the changes discussed above.

1935 Edition (Cohort 1947). At seven years of age a boy was taken away from his home and was put in a camp. There he lived an out-of-door life. He was taught all kinds of exercises, such as running and jumping. Many times he had to go hungry. Often he was whipped severely, but he must never cry out, no matter how great the pain. (pp. 124-125)

1942 Edition (Cohort 1955). At that age every boy had to go to a military camp, and there he had to stay until he was 60. Under such a plan there could be little home life. In camp the boys were drilled for long hours. They were beaten and were made to stand hunger and cold. That was supposed to make them strong. (p. 233)

1951 Edition (Cohort 1962). At the age of seven they left home to live in special camps where all the men, even those who were married, came to eat together. They ate plain food, exercised day after day, and were taught to suffer pain without making a sound. (p.36)

1965 (1968) Edition[7] (Cohort 1972). At seven he was taken from home and placed in a camp. About fifteen boys usually lived together in a camp. Of these the bravest and strongest was chosen captain. Those who did wrong had to take the punishment he set for them. The life of the boys in Spartan camps was very hard. They wore very few clothes and no shoes or hats. After they were twelve, they were allowed only one piece of clothing. (p. 98)

1970 Edition (Cohort 1975). At the age of seven years, boys left their families to live at a school. In school, they played games, developed athletic skills, and had other military training. They had very little training in science or art. The only purpose of education was to prepare boys for a military way of life. (pp. 64-65)

The excerpts from the early textbooks have a keen sharpness and clarity. The readability levels are not appreciably different, yet the excerpts from the later textbooks seem to lose the sharpness of focus and clarity of style by "drawing out" the facts with one short sentence after another.[8]

Also, there seems to be a loss of contrast in describing Spartan life. Could it be that in a sincere attempt to make the past relevant to the modern child, some of the authors of the later editions may have underplayed the differences between past and present--making the textbook perhaps more difficult and less interesting to read and understand and learn from?

In Conclusion

The analyses of the sixth grade history textbooks found a few trends of decreasing challenge from the objective measures. The most definite trend was the consistent increase in pictures over the six cohort years--one and one-half times as many pictures used in the textbooks of the declining as in the stable years.

Some changes were noted in Reading Stage Levels, with higher levels for the two stable years and a decrease in the first two declining years. But there was an increase again in the last two declining years.

Can we say from these various indices that the sixth grade history textbooks have declined in challenge? Perhaps not from the objective measures used here. However, if we take seriously the impressionistic sketches, there appears to be a definite change. The earlier books emphasized information presented in a scholarly manner. In contrast, the more recently published books contain more narrative, more pictures, and in full color. One wonders whether the sixth grade pupils for whom these books are written, are able to sort out the historical facts and interpretations from this seeming abundance.

Is it possible that a more compact, less generously illustrated, history textbook can contribute more to their knowledge and understanding of history? To their linguistic and cognitive growth? To growth of their reading ability? Could it be that with these textbooks--less is more?

CHAPTER 7. ELEVENTH GRADE LITERATURE TEXTBOOKS

The range of books sampled covers the years from 1949 to 1968. Although the original plan called for a book to be sampled for each of the cohort years, we had difficulty in securing some of the books in the limited time available for the study. Therefore, only four of the cohort years are represented--two for the stable years, one for the pivotal year, and one for the declining years.

Because of the inclusive nature of an anthology, containing a vast amount of literature of various types, forms, and periods, we originally attempted to include samples from each of these. We soon found this unworkable since the anthologies contained so many types of literature. We also found from a preliminary analysis of the 1963 textbook that poetry and drama could not be validly checked for readability or reading stage levels. Therefore, they were eliminated from these counts.

Findings

The most striking finding is from the readability analysis, which found great consistency in difficulty over the twenty year period (Table 18). The length of sentences and the percent of unfamiliar words remain almost constant over time.

This lack of change is seen also in other aspects of the anthologies (Table 19). They are relatively similar in total number of reading pages and total number of words. There is a difference, however, in the number

Table 18. Readability of Eleventh Grade Literature Textbooks

Date	1949	1952	1963	1968
Publisher	Publisher C	Publisher C	Publisher C	Publisher G
Dale-Chall Readability Formula				
Average sentence length	23.3	23.6	30.9	24.9
Average Dale score (percent of unfamiliar words)	13.0	13.9	12.9	13.4
Average raw score	6.9	7.0	7.2	7.0
Average corrected grade level	7-8	9-10	9-10	9-10

Table 19. Content Analysis of Literature Textbooks for Eleventh Grade

Date	1949	1963	1968
Publisher	Publisher C	Publisher C	Publisher G
Chall's Reading Stage average	4.0	3.7	3.9
Percent of Stage 2	0%	0%	0%
Percent of Stage 3	14%	41%	24%
Percent of Stage 4	72%	41%	57%
Percent of Stage 5	14%	18%	19%
Narrative Prose			
Total number of selections	39	32	35
*Percent of total book	38%	33%	32%
Expository Prose			
Total number of selections	58	66	60
*Percent of total book	26%	24%	24%
Drama			
Total number of selections	4	2	3
*Percent of total book	8%	6%	8%
Poetry			
Total number of selections	147	150	117
*Percent of total book	19%	23%	20%
Types of prose selections			
Percent narrative	39%	32%	36%
Percent expository	57%	66%	61%
Percent drama	4%	2%	3%

*Not mutually exclusive.

Table 20. Features of the Text and Types of Questions in Eleventh Grade Literature Textbooks

Date	1949	1963	1968
Publisher	Publisher C	Publisher C	Publisher G
Total number of reading pages	749	819	732
Total number of words in book (approximation)	452,163	425,254	461,320
Number of words per reading page.	604	519	630
Percent of picture pages.	15%	30%	22%
Average number of questions per story	6	7	16
*Percent of stories including vocabulary study.	100%	80%	80%
*Percent using glossary	50%	0%	0%
*Percent using basewords.	40%	0%	0%
*Percent using context.	10%	25%	0%
*Percent using definition	0%	75%	100%
*Percent of stories which teach other vocabulary besides content vocabulary.	60%	20%	40%
*Percent of stories suggesting additional reading	60%	60%	20%
Types of questions:	N/31	N/34	N/80
Percent of affective	26%	12%	1%
Percent of who, what, when, where.	52%	29%	30%
Percent of why, how, explain, describe	6%	41%	58%
Percent of compare, contrast	13%	15%	10%
Percent of justify, support.	3%	3%	1%

*Not mutually exclusive.

of pictures, with the lowest percentage in the 1949 book--the one probably used by one of the stable SAT groups. The highest percent of pictures is in the 1963 book, one most likely used by students in the declining SAT years.

With regard to vocabulary study (Table 20), it would seem that the 1949 textbook made the greatest provision for vocabulary study, with 100 percent of the stories containing vocabulary study, compared to 80 percent in the 1963 and 1968 books. Moreover, the 1949 edition had the largest percentage of stories that taught vocabulary other than that presented in the story itself. The 1949 (and the 1963) books suggested additional reading for 60 percent of the stories. This was more than in the 1968 edition, which suggested additional reading for only 20 percent of the stories.

Another change seems to be in the number and type of questions asked following the reading selections (Table 20). The more recent books appear to have more questions. In addition, the quality of the questions asked has changed. In the 1949 book, one fourth of the questions asked were of an affective nature--asking the reader how he or she feels or thinks about the selection. Another half are straightforward, factual questions, asking who, what, where, or when. The 1963 and 1968 books put more emphasis on asking for explanations and/or descriptions, or asking questions of how and why.

Thus it appears that the literature textbooks have been extremely stable in terms of readability, dropping only .3 in raw score on the Dale-Chall formula and not at all on the corrected grade levels.

There are a few signs, however, of greater challenge in the 1949 book --provision for more vocabulary study and for more additional reading and fewer pictures than in the later books. Yet the later books seem to ask students questions that are more demanding.

Which of these contributes more to the kind of abilities needed for a high SAT verbal score?--to the reading of the SAT passages analyzed? It would seem that for at least some of the students, those who have not grown up in a literary atmosphere, a more extensive vocabulary and a greater exposure to literature through reading more than the brief selections in the anthologies might be more "nourishing" than discussing at great depths stories that are on a level of reading generally below that of most eleventh graders.

What of the consistently low readability levels of the books--at the stable years and at the declining ones? They all seem to test at about the ninth grade level. Why ninth grade level for eleventh graders?

To gain some insight into these questions, we took a further look at selections included in the anthologies and their authors. All the anthologies appeared to be similar in content and style. Indeed, the distribution of narrative and expository prose, drama, and poetry were found to be similar within each year analyzed (Table 19).

The formats of the books are also basically similar. There is some variation in the placement of poetry and prose, with some books interspersing them and some books keeping them separate. All the books have sections on modern short stories, modern nonfiction and historical periods of literature from the writings of the Puritans and early colonists and revolutionists through the literature of the expanding west and Civil War, up to the present time. There is some difference in emphasis placed on the literature of the influence and growth of the west, with some of the books keeping this as a separate section and others including it under the section on the modern temperament. Within these periods of historical literature, the selections are relatively similar as to type and style, with only some variation in the choice of authors. Among the authors who appear in all the books analyzed are Jonathan Edwards, William Byrd, Benjamin Franklin, Thomas Jefferson, Washington Irving, Edgar Allan Poe, Ralph Waldo Emerson, Henry David Thoreau, and Mark Twain.

There is a greater variation occurring in the choice of authors and their works in those sections dealing with modern writing--fiction, articles, and essays. Readability measures were obtained for these portions of the anthologies and are presented in Table 21. From the table we note an even greater consistency, with three of the four books having a raw score of 6.6 and all four being on a 7-8 corrected grade level. Thus, the modern writing selections are on an even lower readability level than the total collection which averaged about ninth grade level for the four books.

The Reading Stage Level[9] estimates do not seem to vary much either. When the full anthologies are considered (Table 19), the 1949 book appears to be on a somewhat higher Reading Stage Level. The section on modern writing varies in a similar manner (Table 21). Here, too, as for the total anthologies, it appears as if the Stage Level drops somewhat when the modern literature is considered apart from the remaining selections.

These observations lead one to speculate about the anthologies. Is it that modern writers write on a lower readability level than their predecessors? Or is it that the editors of the anthologies selected those examples of prose that they believed would be of interest to and readable by the students for whom they were selected? Or is it that literary works do not fit the readability measures that were originally developed for expository writing? There is more than a slight suspicion that these may all be part of the answer. Various readability checks of fiction and non-fiction have found that fiction has lower readability scores. Indeed, studies of the readability of best sellers have reported an average of about 7-8 grade readability level. Yet one may still ask whether such literature can provide the kind of exercise that seems to be needed to read the selections included in the SATs analyzed by us which range in difficulty from 11-12 grade level to 13-15 grade level, quite above that of the selections sampled from each of the anthologies.

CHAPTER 8. ELEVENTH GRADE GRAMMAR AND COMPOSITION TEXTBOOKS

The eleventh grade grammar and composition textbooks analyzed were published from 1950 to 1973. The 1969 and 1973 textbooks were published by the same company, and were found to be identical except in the treatment of the discussion of standard and substandard English.

It should be noted that only one of the stable cohorts was represented by an analyzed book, the one published in 1950, representing the 1955 SAT cohort. No textbook in grammar and composition was available for the first stable cohort year of 1947.

Table 21. Readability of Sections of "Modern Writings"* In the Literature Anthologies

Date	1949	1952	1963	1968
Publisher	Publisher C	Publisher C	Publisher C	Publisher G
Dale-Chall Readability Formula				
Average sentence length	20.6	18.4	27.1	20.8
Average Dale score (percent unfamiliar words)	10.4	13.5	10.6	12.5
Average raw score	6.3	6.6	6.6	6.6
Average corrected grade level	7-8	7-8	7-8	7-8
Chall's Reading Stage average	4.0	NOT COMPLETED	3.4	3.9
Percent Stage 3	0%		54%	25%
Percent Stage 4	100%		45%	63%
Percent Stage 5	0%		0%	12%

*Modern short stories, modern articles and essay, and new directions in prose

The same kinds of analyses were made for the eleventh grade grammar and composition textbooks as for the sixth grade. To estimate the degree of challenge, we analyzed the examples of good writing from students and of good writing from literary and historical writing by the Dale-Chall readability formula and by the Chall Reading Stages. The literary examples were also rated on a three point scale of content difficulty and on a three point scale of language complexity. (See Appendix for additional detail on these measures.)

Analyses were also made of general features of the books (length, number of words, number of pictures, and so on) and on the amount and kinds of writing assigned to students.

Findings

The findings should be interpreted as merely suggestive, since only one of the textbooks was available to represent the stable SAT years and the 1965 and 1973 books turned out to be essentially the same.

The most definite trend appears to be in the reading stage levels for both the student writing examples and the good literary examples (Table 22). For both of these there was an increase in reading stage level for the two declining years as compared to the stage levels for the stable years. The same trend was found for the linguistic and cognitive levels of the literary samples.

The trends for the readability of student writing samples (Table 22) are not as consistent, although the Dale-Chall formula seems to show an increase from 1967 to 1973--from a corrected grade level of 7-8 to 11-12. The readability levels for the literary samples and for the exercise samples do not show a consistent trend.

Thus it would appear that few trends could be found, but where there is change over time, some of it seems to be in the direction of more rather than less challenge, as we go from the stable to the declining years.

The analysis of features of the book also found few trends (Table 23). The 1969 and 1973 books were the longest compared to the books representing the stable period. But the decreased size in between makes any observation about a trend questionable.

Proportion of pictures to text seems to have decreased. Indeed, the three books representing the declining years had no pictures, while in the 1950 stable year edition, the pictures were 16 percent of the text pages. The vocabulary exercises showed an increase from 1950 to 1973. All of these indices point to some increase in challenge.

A few measures, such as the percent of exercise pages, may point to a loss of challenge, since there appears to be a steady decrease in the number of exercise pages.

Few patterns were found in the amount of writing required by the exercises in the books (see Table 24). Again, the differences may stem as much or more from preferences of the authors and publishers as from general historical changes. The 1969 and 1973 books have the greatest percentage of underline and one-word answer exercises, and the lowest percentage of exercises requiring extensive writing. The 1967 book requires the greatest amount of writing.

If we compare the 1950 text (stable SATs) with the 1969 and 1970 (declining SATs) and omit the 1967, we can detect a decrease in the more

Table 22. Readability Measures and Chall's Reading Stages and Linguistic and Cognitive Levels for Eleventh Grade Grammar and Composition Textbooks

Date	1950	1967	1969	1973
Publisher	Publisher D	Publisher A	Publisher C*	Publisher C*
Student Writing Examples				
Dale-Chall Readability Formula				
Average sentence length	14.2	14.5	21.2	21.2
Average Dale score (percent unfamiliar words)	13.2	10.8	22.2	22.2
Average raw score	6.4	6.0	8.2	8.2
Average corrected grade	7-8	7-8	11-12	11-12
Chall's Reading Stage average	2.2	3.0	2.8	2.8
Percent of Stage 2	75%	25%	25%	25%
Percent of Stage 3	25%	50%	75%	75%
Percent of Stage 4	0%	25%	0%	0%
Literary writing examples†				
Dale-Chall Readability Formula				
Average sentence length	19.3	25.2	28.0	28.0
Average Dale score (percent unfamiliar words)	17.0	21.8	12.5	12.5
Average raw score	7.3	8.2	7.0	7.0
Average corrected grade level	9-10	11-12	9-10	9-10
Chall's Reading Stage average	3.0	3.2	3.2	3.2
Percent of Stage 2	0%	0%	0%	0%
Percent of Stage 3	100%	75%	75%	75%
Percent of Stage 4	0%	25%	25%	25%
Linguistic level - average	2.0	2.5	2.2	2.2
Cognitive level - average	1.7	2.0	2.2	2.2

(Continued)

Table 22 (Continued)

Date	1950	1967	1969	1973
Publisher	Publisher D	Publisher A	Publisher C*	Publisher C*
Exercise sentences				
Dale-Chall Readability Formula				
Average sentence length	16.5	14.0	17.8	17.8
Average Dale score (percent unfamiliar words)	16.0	10.5	18.5	18.5
Average raw score	7.0	6.0	7.4	7.4
Average corrected grade	9-10	7-8	9-10	9-10

*The 1969 and 1973 editions were similar except for a chapter on standard English usage.
†Note: Only three examples of literary "good writing" could be found for 1950 edition. Most of the examples of literary "good writing" were sentence quotations only.

Table 23. Physical and Other Content Features Eleventh Grade Grammar and Composition Textbooks

Date	1950	1967	1969	1973
Publisher	Publisher D	Publisher A	Publisher C*	Publisher C*
Total number of reading pages	474	419	695	695
Total number of words (approximation)	128,444	110,880	173,448	173,448
Number of words per reading page (approximation).	271	265	250	250
†Percent of picture pages	16%	0%	0%	0%
†Percent of pages of vocabulary	16%	28%	25%	25%
†Percent of pages of exercises.	52%	41%	33%	33%
†Percent of pages of sentences, paragraphs, themes	20%	28%	25%	25%

*The 1969 and 1973 editions were similar except for a chapter on standard English usage.

†Not mutually exclusive.

Table 24. Degree of Writing Required in Eleventh Grade Grammar and Composition Textbooks

Date	1950	1967	1969	1973
Publisher	Publisher D	Publisher A	Publisher C*	Publisher C*
Verbal exercises	15%	2%	13%	13%
Written exercises				
Underline, circle, punctuate, or capitalize for answer	13%	10%	17%	17%
One word answer	19%	23%	37%	37%
One sentence answer	29%	20%	20%	20%
More than one sentence answer	7%	13%	4%	4%
Paragraph	7%	13%	4%	4%
Theme, story, report or letter	10%	19%	5%	5%
Combined types of exercises				
Underline, etc and one word answers	32%	33%	54%	54%
Multiple writing answers-more than one sentence, paragraphs and themes	24%	45%	13%	13%

*The 1969 and 1973 editions were similar except for a chapter on standard English usage.

demanding writing assignments and an increase in the less demanding. Thus, we find in Table 24 a decline from 24 percent to 13 percent of writing that involves more than one sentence, or paragraphs, or themes. Writing exercises asking students to underline or to write one word increase from 32 percent in 1950 to 54 percent in 1969 and 1973.

To conclude, our analyses of the eleventh grade grammar and composition textbooks found few consistent trends, owing partly to the fact that only one textbook was available for the stable SAT years and perhaps partly to the fact that the books for the declining years from the same publisher turned out to be almost identical.

Among the few patterns found, some seemed to confirm the decreasing challenge hypothesis--decrease in number of exercises over time and a decrease in the percent of demanding writing assignments.

However, other trends seemed to support a hypothesis of increasing challenge--an increase in reading stage level and an increase in the linguistic and cognitive levels of the literary samples.

Thus, no definite conclusion about the best-selling eleventh grade grammar and composition books can be drawn at this time.

CHAPTER 9. ELEVENTH GRADE HISTORY TEXTBOOKS

Five eleventh grade history textbooks published between 1961 and 1970 were analyzed[10] (two representing the pivotal year and three copyrighted during the declining years). Unfortunately, the two books representing the stable years were not available for analysis in time to be included in this report. Thus, the main purpose--a comparison between textbooks most probably used during the stable and during the declining years--could not be realized. We present our findings for the declining years only--a period of nine years.

Overall, few trends are evident for features of the books (Table 25). The length of the books and the number of words remain fairly stable. The number of pictures also remains stable for one publisher although the pictures increase by 20 percent for the other publisher. Overall, the pictures seem to take up a major portion of the text, although the extent is somewhat below the 80 percent of pictures in the sixth grade history texts of the same copyright years. Thus, since we had, for the sixth grade textbooks, data for the two stable periods indicating substantially fewer pictures, it is likely that the same will be found for the eleventh grade history textbooks published during the stable years.

The readability levels were also quite stable over the nine year period. The small differences found may in fact reflect the authors' styles, rather than trends. Generally, the books measured at 9-10 or 11-12 corrected grade levels with the two latest books (those published in 1969 and 1970) at the lower end, 9-10 grade levels (Table 26).

The findings for Reading Stages (Table 26) also seem quite stable over the nine year period, with a noticeable increase in the average for the 1970 book. The difference seems more striking when we note the increase in samples rated at Stage 4 and Stage 5. One wonders whether this is a reflection of the emphasis on concepts and discovery learning noted as a primary objective in some of the more recent grade 6 history textbooks.

Table 25. Physical and Content Features of Eleventh Grade History Textbooks

Date	1961	1961	1964	1969	1970
Publisher	Publisher C	Publisher F	Publisher F	Publisher C	Publisher F
SAT cohort year	1962	1962	1967	1972	1975
Total number of reading pages	821	655	655	798	662
Total number of words	331,550	282,072	277,949	380,556	281,688
Average number of words per reading page	624	644	653	682	582
Percent of picture pages	42%	61%	69%	62%	60%
Percent of charts, maps, graphs	9%	16%	16%	10%	10%

Table 26. Dale-Chall Readability Formula Measures and Chall's Reading Stages for Passages in Eleventh Grade History Textbooks

Date	1961	1961	1964	1969	1970
Publisher	Publisher C	Publisher F	Publisher F	Publisher C	Publisher F
SAT cohort year	1962	1962	1967	1972	1975
Dale-Chall Readability Formula					
Average sentence length	20.9	21.5	20.9	20.9	21.9
Average Dale score (percent of unfamiliar words)	19.3	22.5	21.6	19.9	19.9
Average raw score	7.7	8.3	8.1	7.8	7.9
Average corrected grade level	9-10	11-12	11-12	9-10	9-10
Chall's Reading Stage average	3.2	3.5	3.5	3.2	3.9
Percent of Stage 3	80%	50%	80%	80%	25%
Percent of Stage 4	20%	43%	20%	20%	58%
Percent of Stage 5	0%	7%	0%	0%	17%

The findings regarding the cognitive level of questions asked also point to consistency among the five books over the nine year period (Table 27) With the exception of only one of the five books, all have at least 80 percent of their questions on the first level--that is, largely factual. Most of the remaining questions are rated as level 2, (15 percent), with fewer than 10 percent of the questions rated as levels 3 and 4. The various textbooks are also quite consistent with regard to assignments in vocabulary and writing.

In conclusion, it appears that the five history textbooks copyrighted between 1961 and 1970 are quite consistent on our various indices of reading, linguistic, and cognitive difficulty. Overall, they seemed to offer similar challenges to the SAT cohorts who might have used them and who took the SAT during the pivotal and declining years.

Since we cannot make a comparison with textbooks from the stable years, it is difficult to answer the basic question of our inquiry--was there a decline in challenge from the stable to the declining years?

For this we propose one objective measure that may prove significant--the number of pictures. In these eleventh grade history textbooks, the average proportion of pictures is considerable--almost 60 percent. There is some evidence that this proportion was lower for the pivotal year (Table 25).

Thus, the eleventh grade history textbooks may show trends similar to those of the sixth grade history textbooks--a decline in challenge from the stable to the declining years. With an increasing amount of space devoted to pictures, it is important to ask whether they are as essential as currently believed.

The readability scores also raise questions (Table 26). It should be noted that three of the five books are rated at 9-10 grade levels. Only two are at 11-12 and these are at the low end of the category. (The total range in raw scores for all five books is only .6.) One wonders why the levels are so close? It would be instructive to analyze the textbooks representing the stable years. Would the readability scores be higher than for those representing the declining years?

We suspect that the general lack of trends in our measures may be explained by the fact that we have analyzed only that part of the trend that represents a plateau in the general decline of challenge. This is indicated by the stable and rather high proportion of pictures, the rather stable and generally low readability levels, and the consistent and rather low cognitive levels of the questions.

At the same time, there seem to be no signs of increased challenge over the nine years. The only evident increase was in Reading Stage level for the 1970 textbook.

We conclude, therefore, that although we do not at this time have the data to determine whether the extent of challenge decreased during the declining relative to the stable years, there are several indications that additional analyses might be worthwhile.

As further evidence that the hypothesis of decreased challenge during the declining years should not be rejected, we present the following impressions gained from studying the five eleventh grade history textbooks of the declining years.

Table 27. Questions and Followup Activities in Eleventh Grade History Textbooks

Date	1961	1961	1964	1969	1970
Publisher	Publisher C	Publisher F	Publisher F	Publisher C	Publisher F
Average number of questions asked per sample chapter.	17.5	47.5	47.5	8.5	41
Percent of number 1* questions. . . .	82%	85%	85%	56%	86%
Percent of number 2* questions. . . .	16%	12%	12%	29%	7%
Percent of number 3* questions. . . .	1%	3%	3%	9%	6%
Percent of number 4* questions. . . .	1%	0%	0%	6%	1%
Is vocabulary taught?	Yes	Yes	Yes	Yes	Yes
Is further reading suggested?	Yes	Yes	Yes	Yes	Yes
Is further writing suggested?	Yes	Not specified but inferred	Yes independent activities	No	No

*Number 1 questions ask who, what, when, where, list
Number 2 questions ask why, how, explain, describe
Number 3 questions ask compare, contrast
Number 4 questions ask justify, support
(See appendix B for further discussion.)

Impressionistic Views

All the books are similar in physical features and content. They are attractively bound, comparable in size, format and print. One of the earlier publications is slightly smaller, allowing less white space on the pages. Essentially, though, the books seem quite similar. The number of illustrated pages is fairly consistent. Although one of the earlier books contains 42 percent picture pages in relation to total reading pages, the other books range from 60 percent to 69 percent. The illustrations are photographs and reprints from historically significant paintings and art objects. One publisher included drawings made expressly for the book. The illustrations appear to be predominantly black and white rather than full color.

All the books contain little vignettes or capsule studies on historical persons or events. Historical documents such as the Declaration of Independence and Supreme Court Decisions are also included. Suggested additional readings for students are found in each book and the latest publications contain bibliographies for teachers.

The writing style of the books is narrative. One author notes that this style is used to enhance the presentation of history as a good story. The idea of history as a story is further emphasized in the use of suspense to hold the reader's interest.

The more recent authors refer to the need for innovative teaching, quoting such authorities as Edwin Fenton, F. W. Maitland, and Jerome Bruner.

The impression one gets in examining the purposes and rationales of the books is that more attention is given to presenting a general broad overview rather than a basic core of knowledge. Indeed, one publisher states specifically that his book presents history from such a broad perspective.

In summary, these eleventh grade history textbooks seem similar. They appear to give a broad direction to the study of history with an emphasis on general understanding rather than on specific factual knowledge.

CHAPTER 10. SUMMARY, CONCLUSIONS, RECOMMENDATIONS

At the request of the Advisory Panel on Score Decline, we undertook an exploratory study of the relationship between textbooks and SAT verbal scores. Essentially we sought to find out whether the textbooks had changed during the score declining years (1967 to 1975), and if so, whether these changes were related to the SAT verbal score declines.

To study the relationship we analyzed the reading selections from six SATs, two from the stable years (1947 and 1955), one from a pivotal year (1962), and three from the declining years (1967, 1972, 1975). To study the nature of the textbooks during this period of nearly 30 years, we analyzed the most widely used textbooks within selected curriculum areas, grades, and publication dates. The disciplines and grades sampled were: reading, grades 1 and 6; grammar and composition, grades 6 and 11; history, grades 6 and 11; literature, grade 11. The publication dates selected were those most likely used by pupils who took the SATs during the six years selected for the study: 1947, 1955, 1962, 1967, 1972, and 1975. The research design simulated a series of six longitudinal studies, with six hypothetical cohorts who entered grade 1 during specified years, were exposed to specified best-selling textbooks while they were in grades 1 to 11, and took the SATs during the six specified dates noted above.

The various textbooks (35 textbooks and approximately 20 workbooks and teachers' guidebooks) and SAT reading passages were analyzed using various indices of difficulty and challenge--the Dale-Chall Readability Formula for level of reading comprehension difficulty, Chall's Reading Stages for level of linguistic and cognitive maturity, a rating scale for assessing question difficulty adapted from Bloom's Taxonomy, and so on. The textbooks were also analyzed for various physical and content features such as length of book (approximate number of pages), ratio of pictures to text, extent of narrative in relation to expository prose, use of original literature, and the like.

Where the necessary manuals and workbooks were available, we analyzed the kinds of things pupils are asked to do in their workbooks and teachers are asked to do in the teachers' guidebooks.

Not all of the textbooks and SAT passages were analyzed by all of the measures. Some of the indices were standard measures such as the Dale-Chall Formula, and various vocabulary measures used (used particularly for the early readers). Other measures, such as the Chall Reading Stages and the Question Complexity Rating Scale have theoretic validity, but have not been subjected to the usual tests of reliability and validity. Still others, the measures of various physical features and the emphasis given to various activities in the workbooks and teachers' guidebooks, for example, had been used previously by the principal investigator (Chall, 1967).

This was an exploratory study with many limitations. It was especially limited by the use of only a small sample of textbooks (35 plus 20 workbooks and manuals). It would have been helpful to sample more books from more grades. Another limitation was the unavailability of several of the textbooks used by the stable cohorts of 1947 and 1955. The publishers no longer had copies in their archives.

Summary of Findings

In spite of these limitations, we present our findings which should be viewed as highly suggestive, but also as deserving serious consideration and continued study.

1. There seems to be rather compelling evidence that a positive relationship exists between the level of challenge of textbooks used by pupils and the scores they achieve on the SAT-verbal tests. The relationship seems to be stronger for the elementary as compared to the high school years, and for the earlier as compared to the later elementary grades.

Thus, when we map the most widely used first grade reading programs over a 30 year period (representing the SAT cohorts from 1947 to 1975), we find that those cohorts that were in first grade when the reading programs were more challenging did better on the SAT 10 years later. Those who were in first grade when the reading programs had become less challenging, had lower SAT scores 10 years later.

The evidence for the sixth and eleventh grades was not as clearcut as for the first grade. But here, too, there were signs that various measures of lessening of challenge were related to the declining SAT scores. The sixth grade reading textbooks, for example, had a lower percent of original selections from literature, more stories written specifically for the readers, a lower ratio of expository to narrative prose--all indices of a decrease in challenge--during the years preceding the decline in SAT scores. For the sixth grade history books there were also signs of a decrease in challenge during the declining score years as compared to the stable years. In general, the sixth grade history textbooks in use during the declining years seemed almost to change their purpose. From books that were meant to be read, they seem to have become encyclopedic magazines intended to be used for browsing or for reference.

The grammar and composition books present a somewhat similar picture. Although some of the measures show no appreciable change from the stable to the declining years, there are signs suggesting a decline in challenge. Perhaps one of the most important findings is the general low demands on pupils to write, during even the stable years and during the declining score years.

The findings for the eleventh grade textbooks are quite similar to those for grade 6. Some of the indices also reveal a possibility of a relationship between the measures of textbook challenge and SAT scores.

The issue of level of reading difficulty will be discussed later after findings on the SAT selections are discussed.

Why were the relationships more definite for grade 1? And less clear for grades 6 and 11? Several explanations come to mind. One is that all the first grade textbooks--from 1930 to 1962--were from one publisher and from one discipline--reading. The sixth and eleventh grade textbooks covered several disciplines and several publishers, and for some of the disciplines, the earlier textbooks representing the stable years were not available for analysis.

One might also explain the difference by a developmental scheme of reading. According to Chall's scheme of reading development, Stage 1 (which includes most children in grade 1) usually requires direct instruction since most children cannot generate their own rules of correspondence of letters and sounds (Chall, 1976). The children are thus more dependent on their teachers and on their textbooks than in later grades. Once the children have mastered the basic decoding skills and can read simple books on their own, much of their learning can be self generated. In the sixth grade, good readers who have access to a library probably read more words, on a higher readability level, in a month than they read all year in a textbook. Even if their textbooks offer limited challenge, their own free reading can make up for some of the difference.

2. Our readability analyses of the SAT reading passages revealed a decrease in difficulty from the two stable years (1947 and 1955), through the pivotal year (1962), to the declining years (1967, 1972, 1975). Indeed the SAT reading passages of the last declining year studied, 1975, had the lowest readability scores of all years analyzed.

The question analysis revealed an opposite trend. The questions on the later tests seemed to require higher levels of thought. Thus, it appears that as the passages became easier the questions became harder. The exception is in 1975. Here both the passages and the questions seemed to be easier.

3. Our analyses also revealed a possible gap between the difficulty of the SAT passages and the difficulty of the eleventh grade textbooks. The SAT passages were the most difficult of any of the materials analyzed--more difficult than any of the textbooks. Two of the SATs were on a 13-15 grade level and four on an 11-12 grade level.

The history, literature, and grammar and composition eleventh grade textbooks had an overall readability level of 9-10 grade level. Although the Dale-Chall grade levels cannot be used as standards of difficulty appropriate for the different grades, it is still valid to ask what effect such a gap may have. Indeed, if our declining challenge hypothesis has some merit, then it might be logical to ask: Can pupils who are not exposed in their classrooms during their eleven years of schooling to reading matter as difficult as the SAT reading selections be expected to do well on them? Would not they have a better chance if they were given such practice in their various reading assignments?

Of course, by the eleventh and twelfth grades, pupils bound for college are reading on their own (or should be) on higher and on lower levels than the textbooks which are usually written to suit most of them on that grade, including the noncollege bound. But it would seem that the gap between the reading difficulty of their textbooks and the SAT selections would present a greater problem to the new college bound students--those who a decade ago would not enter college. Thus it would appear that the gap between the challenge of textbooks and SAT selections might be greater for those students who do not come with the traditional, more academic background.

It is well to clarify our concern with success on the SAT reading passages. What does success on these difficult selections mean, besides a higher SAT verbal score? the SAT reading selections (and questions) seem to represent the kind of reading required of students in most college work requiring serious reading (Perry, 1968; Chall, 1976). In that sense, can the SAT passages serve also as criteria for reaching that stage of reading maturity necessary for learning in college from textbooks and other printed sources?

4. There are signs in our data of a return to a greater challenge in the textbooks, particularly at the elementary level. Thus, the first grade readers began to show some signs of increasing difficulty in 1962. In 1971 the signs were more definite.

A trend toward greater challenge is found also in the 1965 edition of the sixth grade reader. Many of the indices point to a more difficult reader.

The return of a greater challenge in the reading programs, particularly for grade 1, already seems to have some observable consequences. Could it be that these more demanding reading programs are related to the fact that the early grades, 1 to 4, have not shown the recent decline in achievement scores? (Harnischfeger and Wiley, 1976.) Indeed, increasingly one hears that reading achievement scores for these grades have improved. Thus, a recent announcement from the National Assessment of Educational Progress has reported that the fourth grade reading assessment found the pupils in 1975 achieving better than the fourth graders of five years ago.

If the difficulty levels of the textbooks for the upper elementary grades and high school increase as well, a tenable hypothesis might well be that SAT scores may begin to increase in the early 1980s. Perhaps in 1981? 1982?

At the same time, it would be well to be on the alert for possible negative effects of "more difficult" textbooks on the students who are not college bound and who do not take SATs or other college entrance examinations.

5. We found what appeared to be a particularly low level of challenge in writing. Generally, the assignments in the reading, history, and literature textbooks asked mostly for underlining, circling and filling in of single words. Few assignments asked students to write a paragraph, story, letter, or theme. Even the grammar and composition textbooks, ostensibly designed to teach writing, seemed to spend little time on it.

Interpretations and recommendations

Our examination of textbooks and SAT scores suggests that an essential factor in their relationship is proper challenge. Additional research is of course needed to confirm or disconfirm our findings and hypotheses. However, it is interesting to note that the concept of proper match and proper challenge has been a theoretical and empirical concern of the most eminent psychologists and educators--developmental and educational psychologists and researchers. E. L. Thorndike, I. Lorge, A.I. Gates, J. Piaget, E. Dale, for example.

We need a reevaluation again of what constitutes proper challenge for the children of today, their teachers, and their schools. With universal availability of TV and books, better food, better care, and parents who continue to be more highly educated, it appears that some of the prescriptions about what constitutes proper challenge and what constitutes frustration need rethinking. Such reevaluation would be of help to educators and publishers who are often perplexed by the criteria they think others expect them to use. A reevaluation should also help classroom and remedial teachers who seem to accept the generalization that the easier the book, the more successful the child, the more effective the learning.

We need also to rethink the place of content area textbooks. At what level of readability should they be? Here, too, there has for at least the past 50 years been an assumption that the easier the book the better, provided that enough subject matter is included. If our hypothesis has merit, then it would seem that such books need to fulfill yet another purpose in addition to conveying information. They need to teach the process of getting such information from books--and should provide the practice needed to reach, ultimately, the level of maturity assumed by the SAT reading selections and by college reading requirements.

A Concluding Note:

We should like to point out that our study of textbooks does not rule out the possibility of other factors influencing SAT scores--a general decline in reading and particularly the reading of classics and other challenging literature, the increase in enrollment in the "softer" electives with a loss in enrollment in difficult English courses. These may, in turn, have been reactions to the time-consuming viewing of television. Nor do we rule out the possibility that all of these factors are related to the challenge of textbooks and to SAT verbal scores. Our guess is that all are probably interrelated.

In such cases, the task of educational reform seems extremely difficult and almost impossible. Yet our findings, although far from being definitive, do suggest a way to effect a change. The school textbook, although in modern schoolrooms a rather prosaic learning tool, seems, according to our study, to exert an important force on pupil's linguistic skills. The recent increase in reading achievement of children in Grades 1 to 4 following an increase in the challenge of their reading textbooks suggests that attention to optimal challenge of textbooks can go far toward increasing the reading ability, and hence the verbal SAT scores of those who take the test.

It should be recognized that while our findings suggest that the current textbooks may not be sufficiently challenging (particularly those for the high school), it cannot be concluded that all textbooks now published are too easy. For such a statement, more systematic research will need to be undertaken if educational research is to be more than reactive. Earlier recommendations for greater ease of textbooks, without sufficient research on standards for optimal difficulty, probably contributed to the present situation. Therefore, merely increasing the level of challenge, without supportive research may also fail to produce the desired results. Textbooks appear to be so important in the learning of students that full knowledge of optimal challenge is a necessity. Finally, since textbooks appear to have a significant influence on verbal learning, they should be considered as possible direct means of educational innovation and improvement.

NOTES

1. Although the SAT scores covered those who took the examination in the junior and senior years, the textbooks analyzed were from the eleventh rather than the twelfth grade. This was done to permit the textbooks to have a greater impact on the SAT scores, particularly since a substantial proportion of students take the SAT at the end of the eleventh grade or early in the twelfth grade.
2. Author's note: The designation "pivotal year," is given to those years that seem to mark a turning point in score changes. Prior to these years, the scores either increased or remained stable. Following those years, the scores started a period of decline.
3. The mean SAT scores we used for the selection of the stable and declining years came from the College Entrance Examination Board's "Table I. Means, Standard Deviations, and Number of Cases for all Candidates Taking the SAT, by Year 1956-57 through 1974-75" and "Table 1. Candidate Performance on College Board Tests, 1941-59."
4. The cooperation received from the publishers was extraordinary. Not one refused to send us their materials, and some of it was nearly 40 years old. When we could not obtain a book it was because none was available in the publishers' archives. Indeed, when only one copy was

available and we could not travel to the publishers' library to analyze it, a photographic copy was sent to us. To date, no bill has been received. All the publishers were greatly interested in the study and expressed their desire to receive copies of the report when available.

5. The textbook copyrighted 1967 based on transformational grammar has a lower percentage of this type of writing and a higher percentage of the writing considered more demanding.

6. In response to our request for the 1965 textbook, a 1968 edition was received with thepublisher's statement that these books were the same. Therefore, the 1968 edition was analyzed as the representative book, used by the 1972 cohort, for the sixth grade.

7. For explanation of two dates see footnote to Table 15.

8. This is comparable to our findings for sixth grade readers where although the Dale-Chall scores remain essentially the same, the percent of unfamiliar words increases while the average sentence length decreases (Table 6).

9. For the Reading Stage analysis, each selection was looked at as a whole. Determination of Reading Stages in literary fiction was difficult. We considered those stories which dealt with conflict on a level deeper than the surface plot as being at least a Stage 4.

10. Because it was not clear which were the best-sellers for these years, two books published in 1961 and two books--one published in 1969, and one in 1970--by different publishers, were analyzed.

APPENDIX A

Table I. Basic Data on Dale-Chall Readability Measures, Chall Reading Stages, Harvard Readability Assessment Scale, and Analyses of Questions for Each Reading Passage in SAT 1947 (Form VSA 3)

Passage number	1	2	3	4	5	6	7	8	Average
Average sentence length	29	30	19	41	26	34	31	40	31.3
Dale score (percent unfamiliar words)	20	17	30	26	35	21	22	22	24.1
Raw score	8.23	7.81	9.32	9.74	10,45	8.64	8.65	9.40	9.03
Corrected grade level	11-12	9-10	13-15	13-15	16+	11-12	11-12	13-15	13-15
Chall Reading stage level of Passage*	4	5	5	5	4	4	5	5	4.6
Harvard Readability Assessment Scale†	9	10	9	10	9	9	10	10	9.5
Number of words per passage . .	258	279	354	293	205	206	218	279	261.5
Number of questions per passage.	5	6	6	7	6	5	5	5	5.6
Average degree of knowledge required by questions** . . .	1.8	2.3	2.2	3.0	2.0	1.8	3,6	3.4	2.5

*Chall Reading Stages range in increasing levels of maturity from 1 to 5. (See appendices for description.)

†Harvard Readability Assessment Scale ranges from 1-10 for passages of increasing difficulty, equivalent to a range from Grade 1 to Grade 16+. (See appendices for description.)

**Average degrees of knowledge required by the questions range in increasing difficulty from 1 to 4. (See appendices for description.)

Table II. Basic Data on Dale-Chall Readability Measures, Chall Reading Stages, Harvard Readability Assessment Scale, and Analyses of Questions for Each Reading Passage in SAT 1955 (Form DSA 2)

Passage number	1	2	3	4	5	6	7	8	9	Average
Average sentence length.	35	18	26	28	25	29	28	25	44	28.7
Dale score (percent unfamiliar words) . .	23	29	25	30	28	26	32	32	25	27.8
Raw score	9.00	9.11	8.87	9.76	9.30	9.18	10.08	9.93	9.77	9.44
Corrected grade level. .	13-15	13-15	11-12	13-15	13-15	13-15	16+	13-15	13-15	13-15
Chall Reading Stage Level of passage* . .	4	3	3	4	4	5	4	5	5	4.1
Harvard Readability Assessment Scale† . .	9	8	7	9	8	10	8	10	10	8.8
Number of words per passage	315	267	211	262	242	293	225	294	241	261.1
Number of questions per passage	5	7	5	6	5	5	4	6	4	5.2
Average degree of knowledge required by questions**. . . .	2.0	2.4	2.2	2.8	2.8	3.4	3.0	2.3	3.0	2.7

*Chall Reading Stages range in increasing levels of maturity from 1 to 5. (See appendices for description.)

†Harvard Readability Assessment Scale ranges from 1-10 for passages of increasing difficulty equivalent to a range from Grade 1 to Grade 16+. (See appendices for description.)

**Average degrees of knowledge required by the questions range in increasing difficulty from 1 to 4. (See Appendices for description.)

Table III. Basic Data on Dale-Chall Readability Measures, Chall Reading Stages, Harvard Readability Assessment Scale, and Analyses of Questions for Each Reading Passage in SAT 1962 (Form KSA 15)

Passage number	1	2	3	4	5	6	7	Average
Average sentence length	17	32	23	34	26	26	28	26.6
Dale score (percent of unfamiliar words)	5	32	28	31	20	26	31	24.7
Raw score	5.27	10.28	9.20	10.22	8.08	9.03	9.92	8.85
Corrected grade level	5-6	16+	13-15	16+	11-12	13-15	13-15	11-12
Chall Reading Stage Level of passage*	3	4	4	5	4	4	4	4.0
Harvard Readability Assessment Scale†	6	9	9	10	7	8	9	8.3
Number of words per passage	604	95	355	340	408	363	370	362.1
Number of questions per passage	5	5	5	5	5	5	5	5.0
Average degree of knowledge required by questions**	2.0	2.6	3.2	3.2	2.2	2.8	2.4	2.6

*Chall Reading Stages range in increasing levels of maturity from 1 to 5. (See appendices for description.)

†Harvard Readability Assessment Scale ranges from 1-10 for passages of increasing difficulty equivalent to a range from Grade 1 to Grade 16+. (See appendices for description.)

**Average degrees of knowledge required by the questions range in increasing difficulty from 1 to 4. (See appendices for description.)

Table IV. Basic Data on Dale-Chall readability Measures, Chall Reading Stages, Harvard Readability Assessment Scale, and Analyses of Questions for Each Reading Passage in SAT 1967 (Form PSA 13cp)

Passage number	1	2	3	4	5	6	7	8	9	Average
Average sentence length	25	26	12	26	30	22	48	40	24	28.1
Dale score (percent unfamiliar words) . .	23	37	13	25	23	24	19	24	23	23.4
Raw score	8.51	10.77	6.28	8.87	8.76	8.52	9.02	9.41	8.46	8.73
Corrected grade level .	11-12	16+	7-8	11-12	11-12	11-12	13-15	13-15	11-12	11-12
Chall Reading Stage Level of passage*	4	4	4	4	4	4	5	4	4	4.1
Harvard Readability Assessment Scale† . .	9	9	8	9	9	9	9	9	9	8.8
Number of words per passage	417	380	567	362	504	526	426	474	493	461.0
Number of questions per passage	5	5	5	5	5	5	5	5	5	5.0
Average degree of knowledge required by questions**	3.0	3.2	2.6	3.4	2.4	3.8	3.2	2.4	2.8	3.0

*Chall Reading Stages range in increasing levels of maturity from 1 to 5. (See appendices for description.)

†Harvard Readability Assessment Scale ranges from 1-10 for passages of increasing difficulty equivalent to a range from Grade 1 to Grade 16+. (See appendices for description.)

**Average degrees of knowledge required by the questions range in increasing difficulty from 1 to 4. (See appendices for description.)

Table V. Basic Data on Dale-Chall Readability Measures, Chall Reading Stages, Harvard Readability Assessment Scale, and Analyses of Questions for Each Reading Passage in SAT 1972 (Form USA 25)

Passage number	1	2	3	4	5	6	7	8	9	Average
Average sentence length	30	22	23	17	42	26	21	26	45	28.0
Dale score (percent unfamiliar words) . .	26	26	13	24	24	31	37	16	25	24.6
Raw score	9.23	8.83	6.83	8.27	9.49	9.82	10.50	7.45	9.86	8.91
Corrected grade level .	13-15	11-12	7-8	11-12	13-15	13-15	16+	9-10	13-15	11-12
Chall Reading Stage Level of passage* . .	4	4	3	5	5	5	5	4	4	4.8
Harvard Readability Assessment Scale† . .	9	9	6	9	10	10	10	9	9	9.0
Number of words per passage	318	441	404	429	408	436	411	384	532	418.1
Number of questions per passage	5	5	5	5	5	5	5	5	5	5.0
Average degree of knowledge required by questions** . . .	2.8	3.4	2.6	3.2	2.6	3.4	3.4	3.2	3.4	3.1

*Chall Reading Stages range in increasing levels of maturity from 1 to 5. (See appendices for description.)

†Harvard Readability Assessment Scale ranges from 1-10 for passages of increasing difficulty equivalent to a range from Grade 1 to Grade 16+. (See appendices for description.)

**Average degrees of knowledge required by the questions range in increasing difficulty from 1 to 4. (See appendices for description.)

Table VI. Basic Data on Dale-Chall Readability Measures, Chall Reading Stages, Harvard Readability Assessment Scale, and Analyses of Questions for Each Reading Passage in SAT 1975 (Form XSA 1020)

Passage number	1	2	3	4	5	6	7	Average
Average sentence length	26	23	24	18	22	24	18	22.1
Dale score (percent unfamiliar words) .	12	28	20	27	27	27	16	22.4
Raw score	6.82	9.20	7.98	8.79	8.99	9.09	7.06	8.27
Chall Reading Stage Level of passage*	4	4	4	4	4	4	4	4.0
Harvard Readability Assessment Scale† .	8	9	8	9	8	8	9	8.4
Number of words per passage	401	424	450	448	420	367	584	442.0
Number of questions per passage	5	5	5	5	5	5	5	5.0
Average degree of knowledge required by questions**.	1.8	2.6	2.8	1.8	2.4	2.8	2.4	2.4

*Chall Reading Stages range in increasing levels of maturity from 1 to 5. (See appendices for description.)

†Harvard Readability Assessment Scale ranges from 1-10 for passages of increasing difficulty equivalent to a range from Grade 1 to Grade 16+. (See appendices for description.)

**Average degrees of knowledge required by the questions range in increasing difficulty from 1 to 4. (See appendices for description.)

APPENDIX B

Question Complexity Rating Scale

This scheme was devised specifically for this study, to judge the questions on the SAT reading passages by measuring and numerically rating their degree of difficulty.

It is a judgmental scale, based primarily on the definition of knowledge as outlined by Bloom in his Taxonomy of Educational Objectives (1956) and including ideas on comprehension levels defined by Robert Ruddell (1974) and Frederick B. Davis (1968). In defining the construction of our scheme, we were strongly influenced by Dr. Edgar Dale's rubrics of "reading the lines; reading between the lines; reading beyond the lines."

Guidelines for application of this measure are:

Level 1 - Knowledge of facts (reading the lines): recall and reiteration of facts; identification of specifics such as dates, events, persons, places.

Level 2 - Knowledge of methods of dealing with specifics (reading the lines and between the lines): understanding organization, form and structure such as main ideas, inference; understanding style such as metaphors, verse, analogies and general scientific writing.

Level 3 - Knowledge of highly technical, specific and descriptive terminology (reading between and beyond the lines): understanding vocabulary of attributes, properties, relations, specificity and precision; knowledge of the multiple meanings of words.

Level 4 - Knowledge of universals and abstractions (beyond the lines): ability to use major schemes, principles, generalizations, theories, structure, and their interrelationships; knowledge of complex analogous relationships.

Questions found in history and literature textbooks were evaluated by applying this scheme to specific words within the questions.

Guidelines for application were:

Level 1 - Questions requiring recall and reiteration of facts, asking what, where, who, when, list, define.

Level 2 - Questions requiring recall and use of fact, asking how, why, describe.

Level 3 - Questions requiring the acknowledgment of more than one point of view, asking compare, contrast.

Level 4 - Questions incorporating the first three levels, requiring the reader to justify, support.

Assessment Schemes for Linguistic and Cognitive Complexity for Grammar and Composition Textbooks

The schemes used for assessing the linguistic and cognitive complexity of reading passages were devised by Glenda Bissex, doctoral student, Harvard Graduate School of Education. These schemes, each using a three point scale, were developed for rating examples of good paragraph writing in grammar and composition textbooks. They are presented below:

Language - from familiar/informal/easy to formal/difficult

1. familiar vocabulary and syntax

2. more formal vocabulary and syntax

3. formal - technical or sophisticated vocabulary, elaborated sentences whose structures are characteristic of written rather than spoken language

Content - from close to distant

1. familiar life experiences presented in a straightforward manner

2. content more removed in time and place, or familiar content distanced by analysis or unusual point of view

3. complex content - highly detailed or highly abstract and analytical

Harvard Reading Assessment Scales - by Jeanne S. Chall, Glenda Bissex, Sue S. Conard, Susan H. Harris, Harvard Graduate School of Education (in press)

This scheme is based on a scale of graded passages developed within three content areas, increasing in difficulty from level 1 to level 10 (grades 1 - 16+). It is a judgmental instrument, offering a rapid, qualitative method of matching material and reader.

The scales were used in the analysis of the SAT reading passages. Initially, they were used in the analysis of eleventh grade history textbooks, but because of limited time and a high correlation with the Dale-Chall Formula, they were discontinued.

Chall's Reading Stages - A Measure of the Reading Maturity Required by the Reading Matter

To assess the levels of processing required by the SAT reading passages and the selected textbooks, we used the descriptions of the five Reading Stages (Chall, 1976) indicated below. The reading selection, story, and/or passage was read in its entirety and the Reading Stage Level assigned was the highest level of processing required for understanding what was read.

Stage 1 - Reading that is still mainly associating written to spoken words (decoding).

Stage 2 - Reading that confirms what is already known. Reading "the familiar" for practice in developing fluency. Energy is still concentrated on words, therefore vocabulary is within meaning vocabulary, although not always recognition vocabulary.

Stage 3 - Reading for new knowledge from one viewpoint: facts, concepts, how to do things. The reader is also learning a process--how to find information in a paragraph, chapter, book, etc. The vocabulary goes beyond the elemental, common experiences, being words that are unfamiliar and are learned from books and school.

Stage 4 - Reading from multiple viewpoints; involving greater depth of treatment, dealing with more than one set of facts, theories, viewpoints.

Stage 5 - Reading from a qualitative, relativistic world view. Reading used as one of many forms of knowing and experiencing. Reading that is selective, the ability to read as little or as much of a book as is necessary for the purpose.

BIBLIOGRAPHY

Bloom, Benjamin S., Editor (and others). Taxonomy of Educational Objectives Handbook I: Cognitive Domain. New York: David McKay Co., Inc., 1956, pp. 201-204.

Chall, Jeanne S. "The Great Debate: Ten Years Later, With a Modest Proposal for Reading Stages" (Keynote Address). In L. Resnick and P. Weaver, Editors Theory and Practice of Early Reading. Highland Park, New Jersey: Lawrence Erlbaum, in press 1976.

Chall, Jeanne S. Learning to Read: The Great Debate. New York: McGraw-Hill, 1967.

Chall, Jeanne S. Readability: An Appraisal of Research and Application. Columbus, Ohio: Ohio State University Press, 1958. Reprinted (1974) by Bowker Publishing Co., Ltd., Epping, Essex, England.

Dale, E., and Chall, J. A Formula for Predicting Readability. Bureau of Educational Research, Ohio State University, Columbus, Ohio, 1948.

Davis, Frederick B. "Research in Comprehension in Reading." Reading Research Quarterly, 1968, 4, pp. 499-545.

Fenton, Edwin. Teaching the New Social Studies in Secondary Schools: An Inductive Approach. New York: Holt, Rinehart & Winston, 1966.

Gates, Arthur I. Interest and Ability in Reading. New York: Macmillan Co., 1930.

Harnischfeger, Annegret, and Wiley, David E. "Achievement Test Scores Drop. So What?" Educational Researcher, 1976, Vol. 5, No. 3, pp. 5-12.

Horn, Ernest. Methods of Instruction in the Social Studies. New York: Charles Scribner's Sons, 1937.

Huey, Edmund Burke. The Psychology and Pedagogy of Reading. New York: Macmillan Co., 1908. Reprinted by MIT Press, 1968.

Klare, George R. The Measurement of Readability. Iowa State University Press, 1963.

Perry, William G., Jr. Forms of Intellectual and Ethical Development in the College Years. New York: Holt, Rinehart & Winston, 1968.

Popp, Helen. "Current Practices in the Teaching of Beginning Reading," in Toward a Literate Society, John B. Carroll and Jeanne S. Chall, Editors. New York: McGraw-Hill Book Company, 1975, pp. 101-146.

Ruddell, Robert B. Reading-Language Instruction: Innovative Practices. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1974.

Summary of Score Changes (in Other Tests)

T. Anne Cleary
Sam A. McCandless
College Board

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

December 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

SUMMARY OF SCORE CHANGES*

Let us now consider the information we have on testing programs other than the SAT. In Table 4 we have a compendium of the data that we have been able to collect about other testing programs. Tests are separated by general subject matter, and the observed changes are indicated by plus or minus signs. An attempt has been made to express the observed changes in terms of standard deviation units per year, with three minuses, for example, indicating 3 percent a year, two minuses 2 percent a year. This was done to avoid confusion among the different score scales on which test results are reported.

Score declines have been observed in so many different groups and so many tested areas that serious attention must be devoted to them. The declines have not been large in any given year, but they have been quite consistent over time, area, and group.

The period around 1965 is a critical one for the interpretation of score changes: it is at that time that almost all the score declines started or became larger. The fact that score declines began to be observed at many different grade levels and different subject areas at the same time is impressive and requires that all hypothesized causal agents must be evaluated in terms of their ability to affect all grades and all tested areas at about the same time.

Score declines generally have steadily increased as the ages of the students have increased. In the early grades of elementary school, scores have generally remained stable or increased; most declines have been observed in the upper grades of secondary school.

The decline of scores on the major college admissions tests (ACT and SAT) has been far greater than the declines observed in samples designed to represent all students in a given grade or age level.

The overall decline in scores on the major college admissions tests (ACT and SAT) have been comparable (about 3 percent of a standard deviation per year) since 1965; in both test programs the declines have been greater for women than for men and greater for verbal than mathematical scores; the two programs differ in that the ACT standard deviations have shown a steady increase since 1965, whereas the SAT standard deviations have remained constant.

Among the various tested areas, the most pronounced declines have been observed in social studies, but these declines can probably be attributed to the increasing difficulty of testing knowledge in an area in which so many different curriculums are available.

Scores on tests measuring language skills have declined more rapidly than scores on tests measuring mathematical skills.

---

*Excerpt from T. Anne Cleary and Sam A. McCandless, "Score Declines and Grade Inflation," pp. 90-107 in Northwest Association of Schools and Colleges, Proceedings, Annual Convention, December 1976. Seattle, Wash.: Northwest Association of Schools and Colleges, 1977.

Table 4. Summary of Score Changes

General Ability

Test	Source	Population	Years	Observed changes	Comments
ACT Composite	Crockett (1976)	ACT college-bound	1965-76	---	The decline is about 3 percent of a standard deviation per year, but is greater for females (4 percent) than males (2 percent).
ITED Composite	University of Iowa (1971)	9, 10, 11, & 12 grade	1957, 1962, 1971	Stable	The 1971 norm studies are based on 35,000 students nationally selected.
ITED Composite	Munday (1976)	Iowa grades 9, 10, 11, 12	1962-66	+	
ITED Composite	Munday (1976)	Iowa grades 9, 10, 11, 12	1967-74	--	The amount of decline increases as grade increases.
SRA Achievement Series-Composite . . .	Raju (1976)	Grade 8	1971, 1975	-	Data based on schools that happened to have administered the test.
SRA Achievement Series-Composite . . .	Raju (1976)	Grade 6	1971, 1975	Stable	Data based on schools that happened to have administered the test.
SRA Achievement Series-Composite . . .	Raju (1976)	Grade 4	1971, 1975	++	Increase of about 2 percent of a standard deviation per year. Data based on schools that happened to have administered the test.

(continued)

Table 4. Summary of Score Changes (continued)

General Ability (continued)

Test	Source	Population	Years	Observed changes	Comments
SRA Assessment Surveys Composite Score . . .	SRA 1974	Grades 2-8	1962, 1971	Stable	
Metropolitan Readiness Test Composite . . .	Harcourt, Brace, Jovanovich	Grade 1	1964, 1974	+	
Language Skills					
ACT English	Crockett (1976)	ACT college-bound	1965-76	---	The decline was greater for females (about 5 percent of a standard deviation per year) than for males (about 3 percent of a standard deviation per year).
SAT-verbal	College Board	SAT college-bound	1956-63	Stable	Although the population size was increasing, apparently more and more able students were being drawn into the SAT pool or students in general were becoming more able.

(continued)

Table 4. Summary of Score Changes (continued)

Language Skills (continued)

Test	Source	Population	Years	Observed changes	Comments
SAT-verbal	College Board	SAT college-bound	1963-67	--	Population size was increasing, growing to two-thirds of the college-bound population. It is quite likely that during this time less able students were drawn into the college-bound pool. The decline is equivalent to about 2 percent of a standard deviation per year.
SAT-verbal	College Board	SAT college-bound	1967-76	----	The most marked SAT-Verbal decline has been observed in this period: about 4 percent of a standard deviation per year. The decline is greater for females than for males.
PSAT/NMSQT-verbal	Jackson and Schrader (1976)	Grade 11	1960, 1966, 1974	Stable	A slight decline observed 1966 and 1974
PSAT/NMSQT-verbal	College Board	PSAT/NMSQT high school juniors	1959-75	Stable	If one considers that the PSAT merger with NMSQT drew more able students into the pool and that the scale drift (Modu and Stern, 1965) slightly inflates scores in recent years, a small decline has been observed.

(continued)

Table 4. Summary of Score Changes (continued)

Language Skills (continued)

Test	Source	Population	Years	Observed changes	Comments
MSAT	Perry and Swanson (1974)	Minnesota high school juniors	1959-66	+	
MSAT	Perry and Swanson (1974)	Minnesota high school juniors	1966-73	--	Decline is about 2 percent of a standard deviation per year. Until 1965, MSAT scores were increasing, but number of candidates was also increasing. A new form was introduced in 1966.
ITED Reading	University of Iowa (1971)	9, 10, 11, & 12 grades	1957, 1962, 1971	Stable	
ITED Language Arts	University of Iowa (1971)	9, 10, 11, & 12 grades	1957, 1962, 1971	Stable	
ITED English	Feldt (1976)	Iowa grades 9, 10, 11, 12	1965-74	--	Decline is about 2 percent of a standard deviation per year.
ITED Literature . . .	Feldt (1976)	Iowa grades 9, 10, 11, 12	1965-74	---	Decline increases as grade increases reach about 3 percent of a standard deviation per year in grade 12.
ITED Vocabulary . . .	Feldt (1976)	Iowa grades 9, 10, 11, 12	1965-74	-	Decline increases as grade increases reach about 2 percent of a standard deviation per year in grade 12.

(continued)

Table 4. Summary of Score Changes (continued)

Language Skills (continued)

Test	Source	Population	Years	Observed changes	Comments
NAEP Functional Literacy	NAEP (1975a)	17-year-old students	1971, 1974	+	
NAEP Writing Mechanics	NAEP (1975c)	17-year-old students	1969, 1974	-	
NAEP Writing Mechanics	NAEP (1975c)	13-year-old students	1969, 1973	-	
NAEP Writing Mechanics	NAEP (1975c)	9-year-old students	1970, 1974	Stable	
TALENT English	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	--	The decline was greater for females than for males.
TALENT Reading Comprehension	Flanagan and Jung (1971)	Grade 10 students	1960, 1970	Stable	
TALENT Reading Comprehension	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	Stable	Although the 1975 sample is relatively small, it may indicate that Flanagan and Jung's 1960-70 comparison missed a peak in the mid-sixties.

(continued)

Table 4. Summary of Score Changes (continued)

Language Skills (continued)

Test	Source	Population	Years	Observed changes	Comments
TALENT Vocabulary . .	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	---	The decline was greater for males than for females.
Iowa Basic Skills Reading	Hoover (1974)	Iowa grades 4, 5, 6, 7, 8	1965-72	--	The amount of decline increases as grade increases. The rate of decline increases with time until about 1970 when the scores stabilize.
Iowa Basic Skills Reading	Hoover (1974)	Iowa grade 3	1965-72	Stable	
SRA Assessment Survey Language Arts	SRA 1974	Grades 2-8	1962, 1971	Stable	
SRA Assessment Survey Reading . . .	SRA 1974	Grades 2-8	1962, 1971	Stable	
Mathematics					
Iowa Basic Skills Math Problems	Hoover (1974)	Iowa grades 4, 5, 6, 7, 8	1965-72	--	The amount of decline increases as grade increases. The rate of decline increases with time until about 1970 when the scores stabilize.

(continued)

Table 4. Summary of Score Changes (continued)

Mathematics (continued)

Test	Source	Population	Years	Observed changes	Comments
TALENT Quantitative Reasoning	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	-	
TALENT Mathematics	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	Stable	
TALENT Computation	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	--	Decline is about 2 percent of a standard deviation per year.
ITED Quantitative Thinking	Feldt (1976)	Iowa grades 9, 10, 11, 12	1965-74	--	Decline increases as grade increases reach slightly more than 2 percent of a standard deviation per year in grade 12.
ACT Mathematics . . .	Crockett (1976)	ACT college-bound	1965-76	---	The decline is at the rate of about 3 percent of a standard deviation per year and is slightly greater for females than males.
SAT-mathematical . .	College Board	SAT college-bound	1956-63	Stable	Although the population size was increasing, apparently more and more able students were being drawn into the SAT pool or students in general were becoming more able.

(continued)

Table 4. Summary of Score Changes (continued)

Mathematics (continued)

Test	Source	Population	Years	Observed changes	Comments
SAT-mathematical . .	College Board	SAT college-bound	1963-67	-	Population size was increasing, growing to two-thirds of the college-bound population. It is quite likely that during this time less able students were drawn into the college-bound pool.
SAT-mathematical . .	College Board	SAT college-bound	1967-76	--	The most marked SAT-Mathematical decline has been observed in this period: about 2 percent of a standard deviation per year.
PSAT/NMSQT-mathematical	Jackson and Schrader (1976)	Grade 11	1960, 1966, 1974	Stable	
PSAT/NMSQT-mathematical	College Board	PSAT/NMSQT high school juniors	1959-75	Stable	If one considers that the PSAT merger with NMSQT drew more able students into the pool and that the scale drift (Modu and Stern, 1965) slightly inflates scores in recent years, a small decline has been observed.
ITED Math	University of Iowa (1971)	9, 10, 11 & 12 grades	1957, 1962, 1971	Stable	

(continued)

Table 4. Summary of Score Changes (continued)

Science

Test	Source	Population	Years	Observed changes	Comments
ACT Natural Science	Crockett (1976)	ACT college-bound	1965-76	Stable	
ITED Science	University of Iowa (1971)	9, 10, 11, & 12 grades	1957, 1962, 1971	Stable	
ITED Natural Sciences	Feldt (1976)	Iowa grades 9, 10, 11, 12	1965-74	-	Decline increases as grade increases reach about 2 percent of a standard deviation per year in grade 12.
NAEP Science	NAEP (1975b)	9-, 13-, 17-year-old students	1969, 1972 or 1970, 1973	-	The decline at each age level is about 2 percentage points which is equivalent to about half a year of learning experience.
SRA Assessment Survey Science . . .	SRA 1974	Grades 4-8	1962, 1971	Stable	

Social Studies

Test	Source	Population	Years	Observed changes	Comments
ACT Social Studies	Crockett (1976)	ACT college-bound	1965-76	----	This is the most rapid decline observed, about 5 percent of a standard deviation per year with women declining more than 6 percent per year.

(continued)

Table 4. Summary of Score Changes (continued)

Social Studies (continued)

Test	Source	Population	Years	Observed changes	Comments
ITED Social Studies	University of Iowa (1971)	9, 10, 11, & 12 grades	1957, 1962, 1971	Stable	
ITED Social Studies	Feldt (1976)	Iowa grades 9, 10, 11, 12	1965-74	---	Decline increases as grade increases reach almost 3 percent of a standard deviation per year in grade 12.
SRA Assessment Survey Social Studies . . .	SRA 1974	Grades 4-8	1962, 1971	Stable	
Special Abilities					
ITED Sources of Information	University of Iowa (1971)	9, 10, 11 & 12 grades	1957, 1962, 1971	Stable	
ITED Sources of Information	Feldt (1976)	Iowa grades 9, 10, 11, 12	1965-74	--	Decline increases as grade increases reach almost 2 percent of a standard deviation per year in grade 12.

(continued)

Table 4. Summary of Score Changes (continued)

Special Abilities

Test	Source	Population	Years	Observed changes	Comments
TALENT Abstract Reasoning	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	++	
TALENT Creativity . .	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	++	The increase was greater for females than males.
TALENT Mechanical Reasoning	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	-	Males decreased slightly while females increased.
TALENT Visualization	Flanagan (1976)	Grade 9, 10, 11 students	1960, 1975	Stable	
Iowa Basic Skills Use of References . .	Hoover (1974)	Iowa grades 3-8	1965-72	Stable	
SRA Assessment Survey Use of Sources . . .	SRA 1974	Grades 4-8	1962, 1971	Stable	

REFERENCES

Breland, H.M., *Grade Inflation and Declining SAT Scores: A Research Viewpoint*. Paper presented at the annual meeting of the American Psychological Association, Washington, D.C., September 1976.

Breland, H.M., *The SAT Score Decline: A Summary of Related Research*. Princeton, N.J.: Educational Testing Service, January 1976.

Crockett, D.S., *Score Decline.* Paper presented at the annual meeting of NACAC, October 1976.

Feldt, L.S., *Personal Communication*. Iowa City, Iowa: University of Iowa, 1976.

Flanagan, J.C., "Changes in School Levels of Achievement: Project TALENT Ten and Fifteen Year Pretests." *Educational Researcher*, 1976, *5*, 9-12.

Flanagan, J.C., and Jung, S.M., *Progress in Education: A Sample Survey (1960-1970)*. Palo Alto, Calif.: American Institute for Research, 1971.

Ford, S., *Grade Inflation and Validity of the SAT*. Princeton, N.J.: Educational Testing Service, 1976. Unpublished manuscript.

Harcourt, Brace, Jovanovich, Inc., *Metropolitan Readiness Tests Research Report. Number 2*. N.Y.: Harcourt, Brace, Jovanovich, 1976.

Hoover, H.D., "A Loss (?) in Test Scores in Iowa," *in* Fasold, J., Hall, M., and Impara, J., *Decline in Standardized Test Scores: A Widespread Phenomenon*. Salem, Ore.: Oregon Department of Education, 1974.

Jackson, R., and Schrader, W.B., *Verbal and Mathematical Ability of High School Juniors in 1974: A Norms Study of the PSAT/NMSQT*. College Entrance Examination Board Research and Development Report 76-77, No. 2. Princeton, N.J.: Educational Testing Service, 1976.

Modu, C.C., and Stern, J., *The Stability of the SAT Score Scale*. College Entrance Examination Board Research and Development Report 74-75, No. 3. Princeton, N.J.: Educational Testing Service, 1975.

Munday, L., *Declining Admissions Test Scores*. ACT Research Report No. 71. Iowa City, Iowa: The American College Testing Program, 1976.

National Assessment of Educational Progress, *Functional Literacy: Basic Reading Performance. Technical Summary*. Denver, Colo.: Education Commission of the States, 1975(a).

National Assessment of Educational Progress, *National Assessment of Science, 1969-1973*. Report 04-S-00. Denver, Colo.: Education Commission of the States, 1975(b).

National Assessment of Educational Progress, *Writing Mechanics, 1969-1974*. Report 05-W-01. Denver, Colo.: Education Commission of the States, 1975(c).

Perry, D., and Swanson, E.O., "Decline in Minnesota college aptitude test scores," *in* Fasold, J., Hall, M., and Impara, J., *Decline in Standardized Test Scores: A Widespread Phenomenon*. Salem, Ore: Oregon Department of Education, 1974.

Rajer, N.S., *Personal Communcation*. Science Research Associates, Inc., 1976.

Science Research Associates, Inc., *SRA Assessment Survey Technical Report*. Chicago, Ill.: Science Research Associates, Inc., 1974.

Stewart, E.E., The Stability of the SAT-Verbal Score Scale. College Entrance Examination Board Research and Development Report 66-7, No. 3. Princeton, N.J.: Educational Testing Service, 1966.

University of Iowa, The Iowa Tests of Educational Development: A Summary of Changes in ITED Norms. Iowa City: University of Iowa, 1971.

A Feasibility Study of the SAT Performance of High-Ability Students from 1960 to 1974 (Valedictorian Study)

Thomas F. Donlon
Gary J. Echternacht
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

February 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $3 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

ABSTRACT

The decline in Scholastic Aptitude Test (SAT) averages for the annual candidate population is made difficult to interpret by the complex possible causes. Changes in tests, in populations, in schools and in social customs may all logically make contributions to the decline. The present study sought to avoid some of these complexities by defining a sample of more able students in schools that were specifically selected for stability over the time interval studied.

The study relied upon the availability of school records and upon extensive cooperation by the schools. An initial unsuccessful effort at data collection led to a reformulated inquiry directed specifically at valedictorians and salutatorians. This redirected inquiry succeeded in securing annual samples of approximately 30-80 students in each even-numbered year from 1960 to 1974, in each of three samples: "experimental" schools, specifically selected for high levels of demographic stability, "comparison" schools, selected for intermediate levels of demographic stability, and a selected group of private schools.

The analysis indicated:

1. Valedictorians and salutatorians of experimental schools specifically selected for demographic stability do not show a score decline, and in fact are best described by linear curves with modest increases.
2. Comparison schools show decline but on a pattern statistically different from that of the total Board population.
3. The private schools show a decline not statistically different from that of the total Board population.
4. The Scholastic Aptitude Test is sufficiently sensitive to ability at the upper levels of scholastic performance to differentiate valedictorians and salutatorians in samples of these sizes.
5. While these select students show sex differences similar to those in the total population, there is no evidence of a sex-rank interaction indicating that it is any harder or easier to become a valedictorian or salutatorian depending on sex.

No straightforward interpretation of any data concerning the score decline is possible. The present study supports the view that the score decline is a consequence of social change and social forces, but cannot establish this as a conclusion.

INTRODUCTION

The four major factors that bear logically on the Scholastic Aptitude Test (SAT) score decline are (1) changes in the content of the tests or in the procedures by which raw scores are related to scaled scores; (2)

Authors' note: Richard Harrison was responsible for data-processing operations and contributed to the design of the analysis. Don Rubin provided extensive guidance and consultation in the design and execution of the analysis.

changes in the populations taking the test, shifts in the number and nature of college aspirants; (3) changes in the schools and their curriculums; and (4) broader changes in life styles -- toward TV, away from reading, etc.

It is impossible to study these factors and the relationships among them clearly. They intersect in too many complex ways. Particularly perplexing are questions concerning the possible changes in the population taking the test. During the critical time period from 1960 to the present there has been a broadening of access to college, so that a given student of today is much more likely to seek entrance than a comparable student in 1960. During the same interval, the postwar population "bulge" has increased the numbers of potential students. These factors combine to make year-to-year comparisons difficult. A common hypothesis concerning the decline builds on this expansion of access to college and attributes the decrease in average scores to the influx of large numbers of less able candidates. Data tend to show, however, that there is an absolute decrease in the incidence of high scoring students. Such a finding runs counter to the influx of "less able candidates." It is difficult to find reliable information upon which to base an analysis.

It seemed appropriate, however, to anticipate greater stability, in terms of population characteristics and perhaps in terms of school factors, for higher-ability students as a group. By the beginning of the period of decline, 1960 or so, most of this group was already oriented toward college, so that in many schools a very large proportion of them took the SAT. In addition, motivation to perform well in school is probably stronger and more uniformly distributed for a select group of more able students than it is in the larger populations defined by the annual candidate groups. There is little reason to suspect a marked year-to-year variation in the quality of these superior students. For these reasons, a sample made up of the academically most able group was expected to exhibit greater stability in SAT scores than the group as a whole.

While able students as a group can be held to show less variations than the total group, it is possible to define a subsample with an even greater potential for stability by taking into account school characteristics. A focus on high ability students is essentially normative: the top 1 percent or 5 percent, for example, of a class. This normative definition will be most stable if the sample avoids schools that experienced a great deal of change during the period in question. For example, schools that were redistricted or are combined with other schools during the critical time period would be expected to have the level of their top students altered. It seemed unlikely that any school was so free of change during the 15-year period 1960-1975 that one could treat the classes as equivalent. But if a school was in the same building, was serving roughly the same size student body, was in a community of somewhat stable size and with a community economy based on one or two major and enduring industries, then greater stability in the incidence of ability levels from year to year might reasonably be anticipated. Restricting a sample to schools with these special selective characteristics was likely to enhance comparability further over time.

The study, then, was an effort to define a more stable subpopulation and to examine score-mean patterns for this more stable group. The study obviously could not resolve the many ambiguities that confront an interpretation of the score decline. However, its focus on the higher-ability student gives it at least a face validity, a greater freedom from such perplexing problems as the possibility of declining motivation to do well on tests or to attend college. If restriction to a more stable subgroup does, in fact, alter the pattern of score decline, then a certain component of the anxiety about the decline could probably be allayed.

At the time it was proposed the study seemed dependent upon the availability of records and was, therefore, described as a feasibility study, seeking to find about 200 schools with data available over the years in question. In addition to the problems of the availability of records, however, there was some uncertainty about the capacity to define a demographically stable group of schools meaningfully. For both sample selection and data collection, therefore, there was clearly a feasibility exploration.

PROCEDURE

The design called for a basic sequence of (1) sample selection, (2) data collection, (3) analysis in terms of feasibility, and (4) analysis in terms of implications for the decline. The first three topics are considered in this section. The empirical findings and their evaluation in terms of implications for the decline are presented in a section devoted to results.

SAMPLE SELECTION

The feasibility of defining a sample with sufficient rigor to permit generalization to larger domains was explored in telephone communication with population experts at the National Center for Education Statistics and at the Bureau of the Census. In each case, while there was interest in the study, there were warnings concerning the wide-ranging population shifts that characterized the United States during the period in question.

These warnings led to the total exclusion from consideration of the block of Southern states, because of the extensive mobility of the black populations during the years in question. Similarly, while pockets of large cities might demonstrate apparent population stability in size, the pattern of neighborhood flux was judged to be so great that no single big city area could meaningfully meet the desired criterion of stability.

In response to these counsels, the study arbitrarily avoided certain regions of the country (states, parts of states) suggested by the experts as demonstrating a lower degree of stability than others. The regions selected were located in 10 states: California, Idaho, Illinois, Indiana, Michigan, Montana, New York, Ohio, Pennsylvania, and Washington. While certain of these states obviously have densely populated areas, the experimental schools were chosen with restrictions that avoided them. Thus, within the regions, each school district was required to conform to the following criteria:

1. population 2,500-10,000
2. a high proportion of people living in the same house in 1970 that they had been living in in 1965
3. at least 50 miles from a major metropolitan area
4. fairly heavy, fairly continuous use of the SAT, as indicated by College Board records.

A few schools were selected that failed to meet these criteria perfectly but were judged to have balancing characteristics. Schools selected by applying the criteria above were called experimental schools. In all, 80 were chosen.

While no control group was possible, it was felt that the interpretation of findings would be made clearer by another sample that varied the criteria. Accordingly, a comparison sample was defined, consisting of schools that were located in the same states, and the same regions within states, as were the experimental schools. There were 80 comparison schools. Their characteristics were:

1. population 10,000-50,000
2. a requirement that a low proportion of people be living in same house 1965-1970
3. a requirement of a "suburban" relationship to some metropolitan area
4. substantial history of students taking the SAT.

In keeping with its role as a feasibility study and in view of strong interest in private schools as subject to or free from decline, a small sample of these schools was also included. Again, an effort at stability was attempted. No obviously changed institution was selected. However, it was recognized that these schools, which function in most cases without a specific geographic base, can be subject to quite dramatic shifts in character without any convenient external indicator of the change. In all, some 70 private schools were selected, of which 35 were single sex, 35 coeducational.

Information concerning the status of schools and school personnel is constantly in need of revision. A telephone inquiry was made to the schools to be contacted, alerting them to the planned communication from the project, establishing the appropriate persons to write to, and verifying the general information about the schools. This initial inquiry led to some sample attrition immediately, as schools were reported to be closed or in other ways inappropriate for the study, e.g., a recent redistricting. The final groups consisted of 78 experimental, 72 comparison, and 66 private schools.

DATA COLLECTION PHASE I

The principal factors that governed the design of instruments for the data collection in this study were the questions of student anonymity, the proper depth of information collection, and the ability of schools to supply information.

The problem of anonymity was generally avoided in the initial mailing that sought simple rank in class for unspecified individuals. However, there was interest in extending the study to a more clinical basis by contacting individual students, and in verifying or supplementing school information with information derived from College Board files. Accordingly, a special announcement in the initial mailing invited the schools to supply the names of the valedictorian and salutatorian, pointing out that this disclosure would be consistent with the policy of most major school systems, and stressing that the information could be seen as falling in the public domain.

The initial planning and review discussions of the study emphasized the desirability of a depth sampling of the schools, with an emphasis on a sizable group of the most able students, and an interest in the performance at other levels within class rank. This viewpoint was reflected in the proposal's description of the information that would be sought:

"The schools would provide information on the top 10 percent of each graduating class, as determined by school records. In addition, there would be a request for information on individuals falling in the percentile ranks from 17.5 to 22.5, from 27.5 to 32.5, and from 37.5 to 42.5. In all, records for approximately 25 percent of a graduating class would be requested. While the full design of an analysis would await some determination of the availability of data, the study would contrast SAT mean score patterns for (a) valedictorians. (b) top 5 percent, (c) top 10 percent, (d) those at the twentieth percentile, (e) those at the thirtieth percentile, and (f) those at the fortieth percentile."

However, it was recognized that the principal barrier to such indepth sampling would be the ability of the schools to cooperate. Balancing the needs of this study both to yield data and to explore the schools' capacity to participate, the 80 schools on each of the experimental and contrast groups were randomly assigned to subgroups receiving requests for 10 percent, 20 percent, or 30 percent information. Forty schools (50 percent of the sample) were designated as 10 percent schools, 20 schools were 20 percent schools, 20 schools were 30 percent schools.

Appendix A includes the materials sent to the schools during this first phase of the data collection. These were the letter of invitation to participate, a brief outline of the study, two forms to be used to characterize the community, and forms for supplying student data. These latter forms varied from school to school depending on the depth of sampling attempted.

Because the study was launched in the spring, the initial inquiry to the schools was sent in May. Response to this inquiry was poor. In July, a telephone survey of approximately 25 percent of the sample was carried out. This survey confirmed the fact that the general pressure of school activity in late spring, coupled with the staff reductions during the summer, had made it difficult for schools to respond.

It also confirmed that even the 10 percent sample was too burdensome a request for most schools. The preparation of the data demanded a knowledge of rank in class in, for example, 1960. This information was simply not available.

DATA COLLECTION PHASE II

In conjunction with Board staff and through contact with panel members, it was learned that a restriction to the top two students in any class, the valedictorian and salutatorian, would be an acceptable modification. This conception of the study as a sample of elite "top few" had been a basic one, winning the project the informal label of "valedictorian study." In addition, there had been the specific inquiry concerning the valedictorians and salutatorians included in the initial mailing to the schools. This inquiry had attempted to secure the names of valedictorians and salutatorians in the service of possible personal followup, but also to provide a check on the adequacy of communications, by confirming the reported scores for these students in SAT files.

The reduced level of inquiry was believed to be more manageable for the schools, and new materials for a mailing were prepared. These new materials essentially reestablished communication, explained the revision in emphasis, and reinvited participation. The need for a second data collection phase put the project much behind schedule. Also, there was some need to adjust the project schedule to conform to the scheduled meetings of the advisory panel. Since many schools did not resume operation until mid-September, and since the autumn meeting of the advisory panel was set for mid-November, it was decided to supplement the second mailing with a program of telephone calls. This was done in early and mid-September.

Appendix B presents copies of the revised letter, a description of the study, and copies of the form that was used in the second mailing. As a result of the renewed effort, responses were obtained from 58 of the experimental schools, from 45 comparison schools, and from 42 private schools. In all, reports on 1,592 students were secured.

ANALYSES IN TERMS OF FEASIBILITY

This study is somewhat reduced in value as a measure of feasibility by two of the characteristics it was given at the beginning. The first of these was an attempt at depth sampling, at winning cooperation from the schools in providing information on students whose rank in class was in lower percentiles, such as the 20th to 25th percentile band. In the planning stage of the study, there was extensive interest in these lower strata as possibly being more reflective of the social changes in the college-going population than samples restricted to valedictorians and salutatorians would be. The intention was to provide information on the impact on candidate means of increasing numbers of lower-strata students over the years. As mentioned earlier, hypotheses about an influx of less able candidates have frequently been advanced as either total or partial explanations of the decline. A depth survey would help to clarify and evaluate these hypotheses.

The second characteristic of the study that diminished its feasibility was its timing late in the school year. This was dictated by the schedule of panel activity and, in part, by the problems of sample definition and materials preparation.

Each of these characteristics worked to the disadvantage of the study. The burden of data collection in the depth model proved too great for the schools. Even at the 10 percent level, school personnel described the task as difficult. Given the special character of the sampling none of these schools were very large. But even for these moderate-sized schools the infeasibility of this in-depth collection is established. With regard to schedule, a better initial response might have been possible had the timing been more convenient for the schools.

The foregoing remarks focus on feasibility as reflected in any problems of winning school participation. The initial inquiry was not successful, yielding only 38 replies to 216 inquiries by July 15, 1976, of which a number were acknowledgments of receipt rather than submission of data. The redefined study enjoyed a much greater response, with a total of 124 schools replying. Not surprisingly, the study seems to have demonstrated the relative accessibility of salutatorian and valedictorian records, in contrast to the records of lower ranked students.

Another aspect of feasibility, partially independent of winning school cooperation, concerns the ability of the schools to supply this type of information even if they are willing to cooperate. The N's in Tables 1 and 2 provide some indication of the severity of these problems. These tables are based on the reports of schools. It can be seen that the information in 1960 was available for only about 35 percent of the cases, and about 70 percent was available in later years. Not unnaturally, some schools had policies against the retention of data for longer periods than the schools themselves found useful. A number of problems reduce such data, such as the identification of certain individual valedictorians or salutatorians as aberrant cases, because of such characteristics as not being a native speaker of English. The cooperation of the schools in pointing out such limits on their data was greatly appreciated.

In spite of some information loss, the record of the present study suggests that it would be possible to carry out a larger study, drawing a more random or representative sample of schools and requesting valedictorian and salutatorian data only. The timing of this inquiry is important, for it seems prudent to avoid an end of the year schedule. In addition, such an investigation should be planned to include extensive telephone contact with the schools. The experience of the present study strongly suggested that specific telephone inquiry greatly increased response.

IMPLICATIONS FOR THE DECLINE

Although this study was conducted as an investigation of the feasibility of obtaining appropriate data, the data collection was large enough to offer interpretable findings on the presence or absence of decline in the three special samples. These findings must be viewed as part of the larger context of the total effort to understand decreasing Scholastic Aptitude Test means. The essential role for the study is as a supplement or companion to other reports, intended to contribute to an overall assessment of the reasonableness of considering complex social changes in the schools as a factor in the phenomena of decline.

The basic questions to be considered in this study would seem to be:

1. Does there appear to have been a decline in the average scores of valedictorians and salutatorians in these special samples over the years from 1960 to 1974?

2. If there is a decline in these average scores, how does this compare with the decline reported for the College Board candidate population? Does it differ in size, or in the year of onset? Are there differences between score patterns observed for the SAT-Mathematical and the SAT-Verbal?

3. Do the trends in the average scores for valedictorians and salutatorians appear to be different among the three samples of experimental, comparison, and private schools? Do these apparent similarities and differences relate logically to what is known about the special characteristics of these samples?

Tables 1 and 2 present a brief summary of the essential data on averages across years. Table 1 shows results for the SAT-V, and Table 2 for SAT-M, for each sample in the study and for these samples combined. In each table, data are also provided on the averages in the same years for the college-going population.

The schools did not actually report the particular SAT administration at which a student was tested, so that the average for a given year for the sample groups may include testings from the junior year, if that was the only SAT taken. The effect of this practice of attributing all sample SAT scores to the year of graduation was felt to be minor and to lead to no systematic bias or error.

Figures 1 and 2 show the information concerning means in graphic form. These curves for the sample groups are, of course, less regular than are those for the total population. In all probability this is a function of sample size, but there are common characteristics for the three sample curves: "peaks" in 1966 and again in 1968 for both SAT-Verbal and SAT-Mathematical, and a "peak" in 1972 for the SAT-Mathematical curves. There is no ready explanation for such peaks, but they are more than random fluctuations, being confirmed as significant variation by the statistical analysis.

Although the year-to-year fluctuations make a visual interpretation difficult, the verbal curves seem to offer more evidence of a decline than the mathematical ones. Among the groups the curves for the experimental group seem to show less of a tendency for decline.

Table 3 presents the standard deviations for the various annual samples. These range from a low of 69 to a high of 104, with the bulk of

Table 1. SAT-Verbal Means, Valedictorians and Salutatorians

Year	Experimental schools		Comparison schools		Private schools		Combined sample		College Board candidates*
	Mean	N	Mean	N	Mean	N	Mean	N	Mean
1960. . .	590	32	593	29	656	54	622	115	477
1962. . .	572	50	647	43	641	64	621	157	473
1964. . .	612	63	651	55	652	67	638	185	475
1966. . .	588	68	634	64	653	74	626	206	471
1968. . .	606	82	669	69	653	72	641	223	466
1970. . .	601	96	645	74	638	75	626	245	460
1972. . .	596	83	634	76	623	75	617	234	450
1974. . .	590	82	618	72	630	73	612	227	440
Total N .		556		482		554		1592	

---

*From Breland, Hunter M., "The SAT Score Decline: A Summary of Related Research." Report to the Advisory Panel on Score Decline, January 1976.

Table 2. SAT-Mathematical Means, Valedictorians and Salutatorians

Year	Experimental schools		Comparison schools		Private schools		Combined sample		College Board candidates*
	Mean	N	Mean	N	Mean	N	Mean	N	Mean
1960. . .	627	32	621	29	654	54	638	115	498
1962. . .	618	50	663	43	640	64	639	157	498
1964. . .	626	63	671	55	679	67	659	185	498
1966. . .	625	68	674	64	663	74	654	206	496
1968. . .	647	82	678	69	662	72	662	223	494
1970. . .	627	96	668	76	649	75	646	247	488
1972. . .	649	81	679	78	652	75	660	234	482
1974. . .	645	82	666	74	650	73	653	229	478
Total N .		554		488		554		1596	

---

*From Breland, Hunter M., "The SAT Score Decline: A Summary of Related Research." Report to the Advisory Panel on Score Decline, January 1976.

Figure 1. Average Scores for Each Sample Group and for College Board Total Population, SAT-V

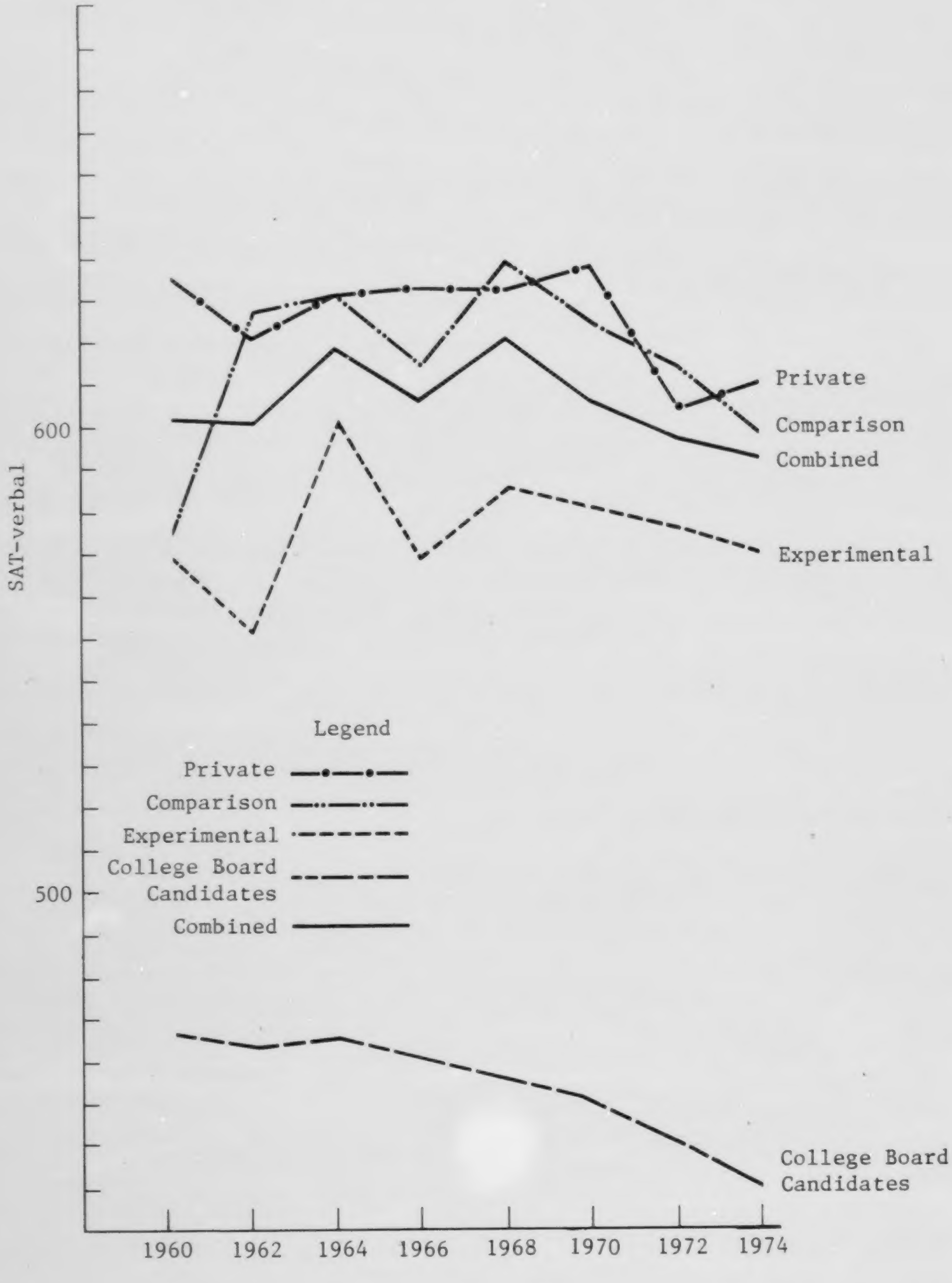
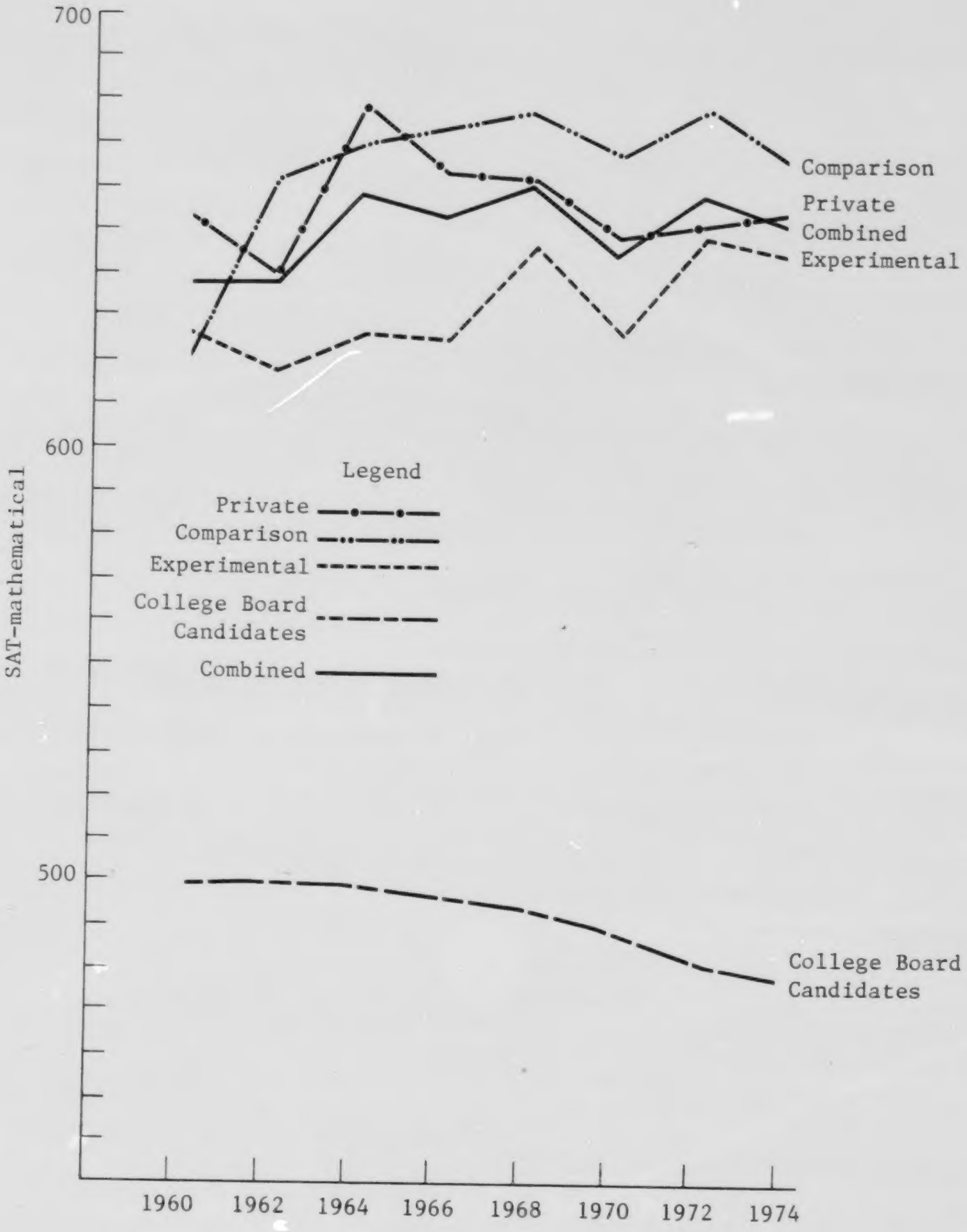


Figure 2. Average Scores for Each Sample Group and for College Board Population, SAT-M



the entries between 70 and 80. The private schools showed the greatest average variation, both for SAT-V and for SAT-M. This could validly reflect greater heterogeneity among the private schools when compared to the public ones. While the public schools are tied to stability of community, the private schools vary in the extent to which they serve any defined local community. The range of values in a number of relevant characteristics (size, location, average income of parents, etc.) was probably greater for the private schools than for the public schools.

Table 3. Standard Deviations of the Sample Groups for SAT-M and SAT-V by Year

	Experimental schools		Comparison schools		Private schools	
	V	M	V	M	V	M
1960	69	90	84	78	87	92
1962	83	75	74	69	91	96
1964	76	85	75	88	80	89
1966	81	92	83	68	90	88
1968	87	76	72	67	88	82
1970	78	70	89	80	91	104
1972	70	71	73	77	82	92
1974	81	82	70	74	88	93

The comparison schools tend to show somewhat less variation within samples on SAT-M, when compared to the experimental schools, and to show somewhat more variation on SAT-V. In view of the sampling procedures, which imposed somewhat more relaxed standards for "stability" on these schools, it might be anticipated that they would constitute a more varied sample.

Perhaps the most striking finding from these tables is the room for measurement at the top of the College Board scale. It is clear that these samples are not representative samples of valedictorians and salutatorians generally. The effort at selecting stable schools would by itself be anticipated to distort selections. The schools picked tended to be smaller and to be located in smaller communities. In view of this smallness, it is probably the case that in the public and private schools here selected it required less average ability to win the competition for top academic honors than might be required in large, urban, elite and science-oriented high schools. Nonetheless, the data show that valedictorians and salutatorians tend to average in the 600s, on both verbal and mathematical tests. They average well above average, and yet there is considerable room for measurement above them in the score scale. However, in connection with the phenomenon of the peaks referred to earlier, there was a need to consider the possible effects of the College Board "800" ceiling. Because the College Board SAT scale is a statistical system, it is not possible to control precisely the scale values of the top scores. These tend to equate the values over 800. In fairness to the top candidates, it is standard policy to report any score of 800 or over, as 800. However, there are occasional test forms for which the maximum scaled score falls below 800, and in certain cases this value is sufficiently close to 800 so that it is allowed to remain. Under certain circumstances, then, such practices of setting arbitrary ceilings could introduce artificial constraints on the data, so that a year with a lower possible top score produced a decline in the average. The wide range of reported scores in this study made this unlikely. To be certain, however, records of the maximum possible scaled scores were examined for the years in question. In no year was there more than a single administration where the phenomenon was observed, and it is unlikely that such a single administration could alter the effects for the total group.

An initial analysis of variance, undertaken with about 80 percent of the data received, indicated significant differences in comparisons of the years, types of schools, the sexes, and the two ranks in class -- valedictorian and salutatorian. The interest in the study lay not simply in the variation among years, but in trends in the data. Accordingly, it was decided in subsequent analyses to test the contribution to prediction of the best fitting curves. Since there were eight years, it was possible to fit curves of increasing complexity from a straight line to a seventh order polynomial equation.

It was further decided to examine the school relationships in terms of two contrasts: the difference between the two public school groups, experimental and comparison, and the difference between the private schools and the combined public schools. Sex and rank were retained as main effects. Table 4 presents the regression weights and the t-values associated with these weights for each major factor in the analysis, for SAT-V and for SAT-M, using the overall sample.

Table 4. Regression Weights and Associated t-Values for Main Factors in the Analyses of Variance

Factor	SAT-verbal		SAT-mathematical	
	Weight	t	Weight	t
Overall mean	625.10	287.20†	651.85	322.25†
Contrast A:				
Experimental vs. comparison	43.52	8.31†	31.39	6.45†
Contrast B:				
Public vs. private	26.39	5.82†	-4.80	1.14
Sex	-1.85	-0.86	24.83	12.36†
Rank				
(Valedictorian/salutatorian)	7.57	3.58†	10.97	5.59†
Year:				
Linear term	-4.23	-0.65	14.83	2.48*
Quadratic term	-21.87	-3.42†	-13.27	-2.23*
X^3 - term	-1.02	-0.16	3.48	0.60
X^4 - term	5.56	0.91	3.41	0.60
X^5 - term	1.75	0.29	-8.92	-1.60
X^6 - term	6.03	1.03	2.43	0.45
X^7 - term	-13.51	-2.34*	-11.21	-2.09*

*$p < .05$
†$p < .01$

The SAT-M score was, on the average, somewhat higher than the SAT-V. This conforms to the pattern for the total candidate population. The gap separating the two means, approximately 25 points, is approximately the same as the 29 points difference for the total population in recent years.

Significant effects were observed for the two contrasts on SAT-V, and for the Experimental versus Comparison contrast on SAT-M. In general, comparison school valedictorians and salutatorians averaged about 40 points higher for SAT-V, about 30 points higher for SAT-M. Since the comparison schools were larger and more urban, this was to be expected. For the other contrast, there was a significant superiority for private school valedictorians and salutatorians on SAT-V, but not on SAT-M.

The differences between the sexes were on the order of the differences observed for the larger population, and in the same direction. Females showed a nonsignificant superiority on SAT-V, males a significant one on SAT-M. In the absence of a knowledge of courses taken, the attribution of these observed differences to causal factors cannot be attempted. Recent years have seen the virtual disappearance of the sex difference in SAT-V. The difference in SAT-M has remained, but the sensitivity of this variable to recency of appropriate course work has long been established (Dyer, 1953; French and Dear, 1958). If nothing else, the data suggest that female valedictorians may retain the general pattern of less mathematical course work which holds for females in the total population.

Rank in class yields significant differences both for SAT-V and SAT-M. This is an extremely interesting finding, since it attests to the sensitivity of these tests even among elite groups. The relevance of the Scholastic Aptitude Test is sometimes challenged for the extreme subgroups at the top and bottom of the candidate distribution. As this study shows, the tests differentiate by small but significant amounts between the approximately 800 valedictorians and the approximately 750 salutatorians. The differentiation is about 8 points in favor of valedictorian on SAT-V, about 11 points on SAT-M. Clearly the tests and the schools seem to be in agreement. Since the distinction that is being made is often a very difficult one, involving individuals of quite similar attainments, the ability of the tests to make the appropriate distinction is significant.

The year effects are described in terms of the significance of each of the seven weights that would be used in a seventh-degree equation to describe the data. Each such weight may be independently tested.

The linear (x) term is significant for the SAT-Mathematical curve, but not for SAT-Verbal. The quadric (x^2) term is significant for both tests. This tends to suggest that for each test the data are well described by a curve that sees the rate of decline increasing over time. Such curves have tended to describe the College Board population as a whole.

A perplexing aspect of the analysis was the significance of the "septic" (x^7) term. The effect of this term on the fit of the curve is difficult to visualize. It is found, however, in both SAT-V and SAT-M. It seems to be enhancing the fit of the equation to the two "peak" years, 1964 and 1968. The finding of such peaks in these special samples was not anticipated. It does not seem to be reflected in the means of the College Board total population. It is possible that it is a phenomenon related to the scaling and equating of the tests although no mechanism by which this could happen was apparent. Since the interest was in the more regular trends in the data, this seventh degree term was ignored in the subsequent analysis.

In addition to the main effects, there are a number of interactions to be tested. Since the years effect is analyzed into seven components, the basic interactions are expanded to 37 second order, 28 third order terms. Table 5 presents those interactions that were significant. Of the small number of significant interactions, only that between the private-public contrast and sex was significant in both analyses. It was considered likely that this was a product of the nature of private schools, with some schools coeducational, others all male, yet others all female.

Table 5. Regression Weights and Associated t-Values for the Significant Interactions

Interaction	Weight	t
SAT-verbal:		
Contrast B x Sex.	-17.68	-3.74†
Year (X^6) x rank.	6.19	2.30*
Year (linear) x contrast A.	-34.01	-2.44*
Year (linear) x contrast B.	-36.96	-2.12*
SAT-mathematical:		
Year (quadratic) x contrast A	-39.88	-2.49*
Sex x contrast B.	-9.96	-2.26*
Year (X^6) x contrast A x rank	28.74	2.00*

*p < .05
†p < .01

It is probably appropriate to comment on some of the nonsignificant interactions. First, it should be noted that there was no significance to the interaction between sex and the experimental versus comparison contrast. The changes over time in these two samples are not statistically different for the sexes for either test. While the sexes differ in performance on the mathematical test, their rate of change over time is, within the limits of probability, the same.

The absence of a significant interaction between sex and rank, for either SAT-V or SAT-M is also interesting. This can be interpreted as indicating that the processes that identify and designate valedictorians and salutatorians are indistinguishable for the sexes from the standpoint of test performance. Put another way, there is no evidence that anything "extra" is required, in terms of SAT performance, to be a valedictorian or salutatorian depending on sex. The test differentiates the ranks and a single differentiation applies regardless of sex.

In the next phase of the analysis, the annual means for each sample were adjusted by subtracting the corresponding annual mean for the total College Board population. The effect of this process is such that if there is a trend in the data of a given sample, and if it matches that of the total Board population, it will be removed. There will be no significant effect for year in the ensuing analysis. At the same time, attention was restricted to a few designated interactions, rather than the complete analysis. Table 6 presents the principal results.

Table 6. Regression Weights and Associated t-Values for the Total Sample: Scores Adjusted for College Board Population Means

Factor	Weight	t
SAT-verbal:		
Overall Mean.	163.46	74.82†
Contrast A: experimental vs. comparison	43.50	8.31†
Contrast B: public vs. private.	28.02	6.11†
Sex	-2.18	-1.00
Rank.	7.55	3.59†
Year (linear term).	36.27	5.58†
Year (quadratic term)	-14.59	-2.32*
Interactions:		
Year(linear) x contrast B.	-30.63	-2.33*
Year (quadratic) x contrast A.	-28.66	-1.87
Year (quadratic) x contrast B.	15.15	1.17
Contrast B x sex	-17.05	3.75†
SAT-mathematical:		
Overall mean.	161.77	79.91†
Contrast A: experimental vs. comparison	31.52	6.47†
Contrast B: public vs. private.	-4.79	-1.13
Sex	24.81	12.35†
Rank.	10.99	5.58†
Year (linear term).	17.05	6.15†
Year (quadratic term)	-4.02	-0.69
Interactions:		
Year (linear) x contrast B	-25.26	-2.06*
Year (quadratic) x contrast A.	-30.21	-2.12*
Year (quadratic) x contrast B.	-1.58	-0.13
Contrast B x sex	-10.02	-2.36*

*p < .05
†p < .01

The results of the analysis in Table 6 essentially show that the adjusted data retains the principal structure of the initial analysis. No main effect is shifted in significance level. The interactions indicate significant effects for the contrasts with some of the terms in the descriptive equation. Private versus public contrast is significantly interactive with the linear term, but not the quadratic term for both tests. The experimental versus comparison contrast shows significant interaction with the quadratic term for the SAT-M, and a borderline value for SAT-V. The interaction of the private versus public contrast and sex continues to be strongly interactive in this analysis.

Since the contrast-year interactions had proved significant, different curves were needed to describe the changes over time for the three samples. Accordingly, a last analysis was made separately for each sample on each test. This analysis retained the use of adjusted scores. Because the emphasis is on comparisons of samples, the results of these separate analysis of variance tables are recast in Table 7.

Table 7. Regression Weights and Associated t-Values for Five Factors in Each Sample

		SAT-verbal		SAT-mathematical	
Factor	Sample	Regression weight	t	Regression weight	t
Overall mean.	experimental	130.80	36.24†	147.33	44.51†
	comparison	172.40	46.44†	175.81	52.86†
	private	182.59	44.25†	159.51	39.91†
Sex.	experimental	3.60	1.05	29.75	9.46†
	comparison	4.09	1.14	26.29	8.16†
	private	-13.71	-3.33†	18.15	4.56†
Rank	experimental	7.91	2.36*	13.13	4.27†
	comparison	7.07	2.00*	11.08	3.51†
	private	7.44	1.88	8.56	2.23*
Linear curve	experimental	41.00	3.77†	40.36	4.05†
	comparison	43.03	3.81†	48.22	4.77†
	private	7.36	0.64	17.46	1.56
Quadratic curve	experimental	-3.45	-0.32	9.19	0.94
	comparison	-39.17	-3.54†	27.09	-2.73†
	private	-0.06	-0.01	-5.95	-0.54

*$p < .05$
†$p < .01$

The comparisons of the overall means in Table 7 establish that the experimental valedictorians were less able than the comparison or private school valedictorians on both SAT-V and SAT-M. The patterns of sex differences were as anticipated, with established differences for SAT-M. Private school female valedictorians and salutatorians were significantly more able than their male counterparts on SAT-V, but this difference was not significant for the other samples.

The findings with respect to rank are also in keeping with previous analyses. There is one nonsignificant value of t for the private school sample, but in view of the overall pattern, this single finding does not alter the general conclusion that valedictorians and salutatorians do differ in all samples.

The major interest lies in the trend lines that describe these "adjusted" scores. This analysis shows that the private schools do not differ significantly from the general trend of College Board decline. None of the four tests of equation weights is significant for the private schools. For the other two samples, there are significant weights, indicating that the trend for the total population cannot describe these special samples. For the comparison schools there are significant weights for both linear and quadratic terms. For the experimental schools, only the linear term is significant.

The general effect of this analysis of adjusted scores is to suggest that the comparison schools show less decline at the outset but a tendency, in the later years, to be declining more rapidly than the total Board population. The experimental schools, however, show positive linear trends over time. The distance between the experimental schools and the total Board population is constantly expanding over the study. This suggests but does not establish that there is no decline for these schools. These adjusted data cannot establish the exact nature of the experimental curve because they are relative data, compiled by subtracting the mean of the total College Board group in each year.

The empirical curves for the unadjusted scores, presented in Figures 1 and 2, contain a number of fluctuations from year-to-year that make it difficult to sense the trends involved. Accordingly, best-fitting curves were calculated for both SAT-V and SAT-M for each sample group and for the total College Board population. On the basis of the earlier analyses, the curves for private and experimental schools were simple linear curves, those for the comparison schools and the total Board population were quadratic. The specific equations used are presented in Table 8. Figures 3 and 4 present the results graphically.

Table 8. Equations of Best-Fitting Curves for Each Sample Group and for Total Board Population

SAT-verbal
Experimental. . . . . $Y = 594.38 + .63\ (\text{year } 1967)$
Comparison. . . . . . $Y = 656.53 + .76\ (\text{year } 1967) - .96\ (\text{year } 1967)^2$
Private . . . . $Y = 643.25 - 1.87\ (\text{year } 1967)$
Total College Board . $Y = 468.63 - 2.52\ (\text{year } 1967) - .22\ (\text{year } 1967)^2$

SAT-mathematical
Experimental. . . . . $Y = 633.00 + 1.82\ (\text{year } 1967)$
Comparison. . . . . . $Y = 678.31 + 2.32\ (\text{year } 1967) - .63\ (\text{year } 1967)^2$
Private . . . . $Y = 656.13 - 0.35\ (\text{year } 1967)$
Total College Board . $Y = 494.50 - 1.50\ (\text{year } 1967) - .14\ (\text{year } 1967)^2$

The curves show that the scores for valedictorians and salutatorians in the experimental school are best described by consistent patterns of increase over the years 1960-1974. This increase is slight, but it indicates that for these schools the pattern is not one of less decline, but of no decline at all. The experience of the comparison sample is somewhat complex in an initial increase and then a decrease. It was the only sample group best described by a curve of complexity higher than linear. The curve slopes up and down quite sharply. An inspection of Figures 1 and 2 indicates that a single year, 1960, is strongly influencing this picture. If this year were removed the general shape might remain the same but the arch of the parabola would be much flatter. The private schools showed a small but steady decline in both SAT-V and SAT-M. This decline is best described by a linear fit, but the pattern is not statistically inconsistent with a random fluctuation from the decline exhibited by the total College Board population.

Figure 3. Best Fitting Curves for Each Sample Group and for College Board Total Population, SAT-V

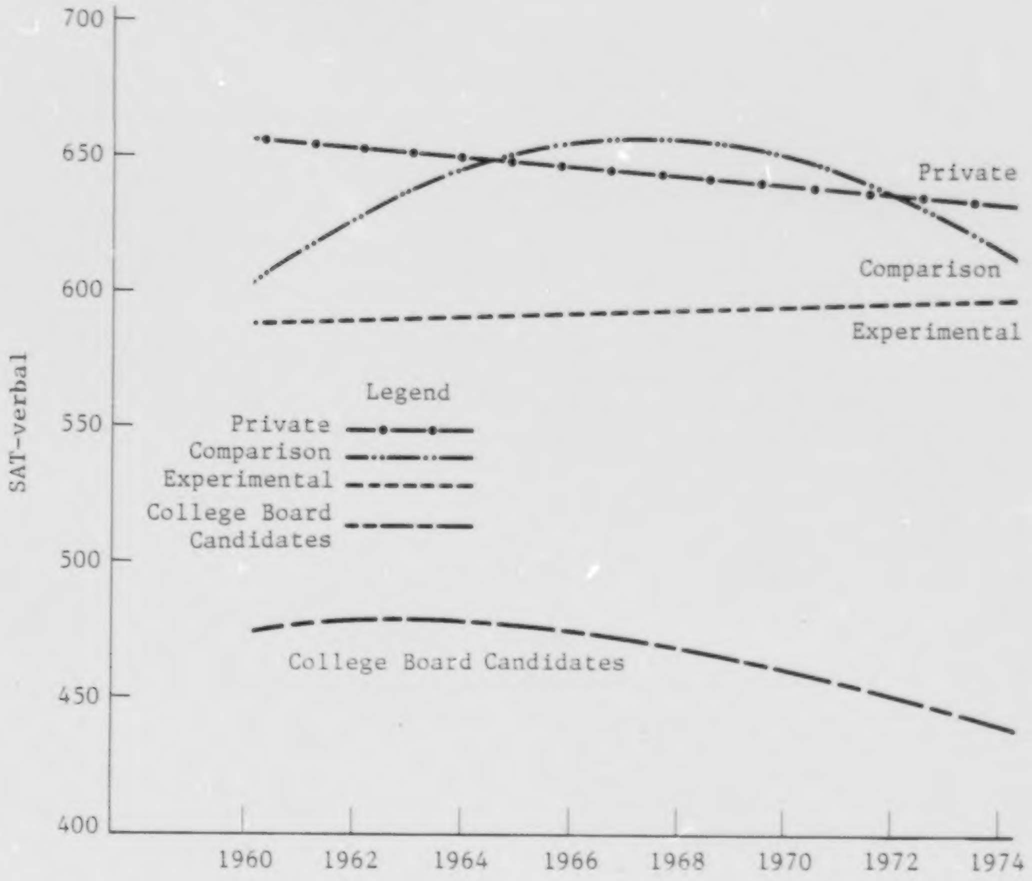
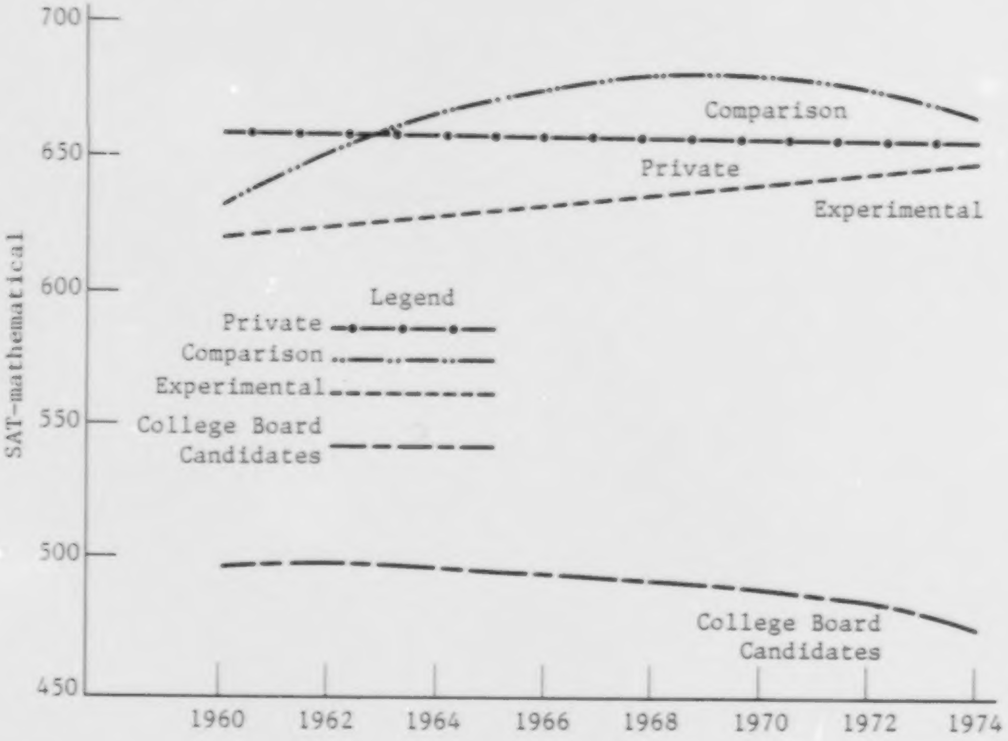


Figure 4. Best Fitting Curves for Each Sample Group and for College Board Total Population, SAT-M



The curve which describes the total College Board population is a quadratic indicating a gradual intensifying of the rate of decline. The curve, being based on the even-numbered years from 1960 to 1974 is of particular interest because it affords a projection, beyond these years to the experience in 1976. The predictions for 1976 are, for SAT-V, 428, and for SAT-M, 469. These may be contrasted with the reported values of 429 and 470. Thus, the observed values correspond so closely to the theoretical as to suggest that the process of the decline was still underway in 1975-76 and producing decrements well described by the curve.

This study was intentionally exploratory and tentative. Nonetheless, it supports conclusions with regard to feasibility and to the presence or absence of decline in the special samples. The conclusions concerning feasibility would be:

1. Enough schools retain sufficient records of test performance to provide for a feasible sampling of valedictorians and salutatorians during 1960-74.

2. Not enough schools retain organized records of test performance to permit an indepth sampling of ranks lower than valedictorian/salutatorian.

The conclusions concerning score decline would be:

1. Valedictorians and salutatorians of experimental schools specifically selected for demographic stability do not show a score decline, and in fact are best described by linear curves with modest increases.

2. Comparison schools show decline but on a pattern statistically significant from that of the total Board population.

3. The specially selected private schools show a decline not statistically different from that of the total Board population.

4. The curve that describes the rate of decline for the total Board population in 1960-1974 quite accurately predicts the means for 1976.

Other conclusions worthy of note are:

1. The Scholastic Aptitude Test is sufficiently sensitive to ability at the upper levels of scholastic performance to differentiate valedictorians and salutatorians in samples of these sizes.

2. While these select students show sex differences similar to those in the total population, there is no evidence of a sex-rank interaction indicating that it is any harder or easier to become a valedictorian or salutatorian depending on sex.

The failure to observe a score decline for valedictorians and salutatorians in specially selected schools lends support to the views that the decline is related to patterns of social change. It must be cautioned, however, and strongly, that this conclusion may be simplistic. The kind of specialized sampling carried out in this study is fraught with hidden consequences. It seems plausible that the communities selected have experienced less social change, and it seems plausible that their valedictorians and salutatorians show less decline because of this reduced change. But the conjecture is only supported, not proven, by the data.

The study indicates that a larger and more comprehensive effort to understand the test-taking patterns of the elite students is feasible and can lead to analyses that improve significantly our understanding of the forces underlying the decline. It is doubtful that there are large numbers of stable communities with large SAT-taking groups to add to the schools considered here. But a random national sample of schools could provide a base from which to undertake demographic analyses of greater precision. For example, do schools that are considered demographically "unstable" show a greater rate of decline than do other schools?

References

Dyer, H.S., "Does coaching help?" College Board Review, No. 19, 1953, pp. 331-335.

French, J.W., and Dear, R.E., "Effect of Coaching on An Aptitude Test." Educational and Psychological Measurement, No. 19, 1959, pp. 319-330.

APPENDIX A

Letter sent to Schools, May 1976
Outline of the Study
Special Request
Community Characteristics, Form A
Educational Characteristics, Form B
Instructions for Score Report, Form C-1, top 10 percent
Instructions for Score Report, Form C-2, top 20 percent
Instructions for Score Report, Form C-3, top 30 percent
Score Report, Form C

EDUCATIONAL TESTING SERVICE PRINCETON, N. J. 08540

Area Code 609
921 - 9000
CABLE-EDUCTESTSVC

Developmental Research Division

May 3, 1976

Declining scholastic aptitude and achievement test scores have received much attention in the last few years. Concerned by the situation, the College Entrance Examination Board has established a panel of prominent educators to review the factors associated with the decline. Under the supervision of this panel, a number of studies are being conducted.

We are writing to request your participation in a study of declining scores for selected populations. As the accompanying material indicates, we are interested in the score patterns of students in schools where there has been little change in the demographic make-up of enrollments. Your school was chosen because census data indicates your community may be considered "stable" in terms of its social and economic characteristics.

As the accompanying outline shows, we have developed three forms for recording information about the characteristics of your community, your school, and a form for recording SAT scores for every other year from 1960 through 1974. We hope that you will be able to answer the questions about community and school characteristics and that an appropriate test officer can complete the score report forms.

We do not require the names of any individual students, only verbal and math scores, rank in class, and sex. We will not reveal the scores of any school to others. However, we will prepare a report for you, providing your school's score averages, those averages achieved by all schools in the sample, and a list of participating schools.

Your cooperation will help us better understand the score decline. If you have any questions about the study, please call either of us collect at (609) 921-9000.

Sincerely,

Thomas F. Donlon
Senior Research Psychologist

Gary J. Echternacht
Research Psychologist

Enclosures

Outline of the Study

Interpreting SAT mean data over a long period of time is made difficult by the shifts in the composition of the group from year to year. As more and more students elect to seek a college education, the groups become more and more diverse. It is possible that a substantial portion of the decline is attributable to this shift in the nature of the group. No research study can fully solve this problem. However, we seek to reduce variability by focussing on students with specific characteristics in terms of academic attainment. Essentially, this focus is achieved by limiting the sample to students with a defined rank in class. Is the decline observed, for example, in samples of students all of whom were valedictorian students standing first in their class? Is it observed for samples of students who rank in the top 10%? If such samples can be identified, it may be assumed that they are less subject to year to year variability than the self-selected groups who attempt College Board Tests. Further, the study attempts to identify schools for which the meaning of rank in class may have been more consistent over the years: schools in communities which have not experienced marked shifts in racial or ethnic composition, in socioeconomic status, in the typical occupations of family breadwinners, etc.

We ask that schools supply data in the specific formats we request. Our project will combine the reports of individual schools into a single coherent report. The initial phase of our study will focus on approximately 60-75 schools. If it is successful, a larger extension of the study will be proposed.

Special Request

Throughout this study we have tried to preserve the anonymity of individual students. However, we have learned through our advisory panel that considerable interest is focussed on the very able valedictorians and salutatorians. It has been suggested that the winning of such honors, by virtue of its publication in the news media, is a matter of public record. We would like to contact a sample of these students, to learn more about them. If school policy permits the release of such data to the College Entrance Examination Board and to the Educational Testing Service, we would like to request the names and addresses of the valedictorians and salutatorians reported in Form C. We hold that this release is consistent with school policy such as the following, formulated by a large metropolitan school district in California:

> Organizations conducting studies for, or on behalf of, educational agencies or institutions to develop, validate, or administer predictive tests, administer student aid programs, and to improve instruction. (Such studies must not permit personal identification of students and their parents by persons other than those conducting studies. Information must be destroyed when no longer needed for the purpose for which it is obtained.)

We would conform to the conduct stipulated in this policy. We would be pleased to attempt to resolve any problems connected with this special request by telephone.

COMMUNITY CHARACTERISTICS FORM A

We are especially interested in the degree to which the socioeconomic character of your school district and community have changed over the past 15 years. Although few areas of our country have been immune to social and economic change during the past 15 years, some communities and school districts have more or less retained their social and economic characteristics. In particular their school enrollments have neither increased or decreased dramatically, the racial composition of the student body has remained the same, the socioeconomic status of the students has remained the same, and the general economic character of the community has remained the same (e.g., predominanatly agricultural products, manufacturing, mining, retail or wholesale trade, transportation, etc.) In the space provided below, please indicate the degree that you would say your school district has remained socially and economically stable over the past 15 years.

Very Unstable	Unstable	Stable	Very Stable
☐	☐	☐	☐

Also, answer the following specific questions.

	Yes	No
1. Has your school district's enrollment increased or decreased by more than 10% over the last 15 years?	☐	☐
2. Has the ethnic or racial composition of your enrollment changed by more than 3% over the last 15 years.	☐	☐
3. Has the socioeconomic status of your enrollment changed substantially over the last 15 years?	☐	☐

4. Please indicate by checking one box below the major economic activity of your community.

Agricultural Products	Manufacturing	Mining	Retail/Wholesale Trade	Transportation
☐	☐	☐	☐	☐

Other - specify ______________________________

____

____

In the space provided below would you please provide a brief description of the social and economic character of your community and how it may have changed over the past 15 years.

EDUCATIONAL CHARACTERISTICS FORM B

We are interested in changes in the schools themselves, particularly as these center on concepts of academic excellence. It is inconceivable that education has not changed extensively in process during the years 1960-75. Our study cannot assume therefore, that the best students of 1960 are equivalent to those of 1974. But we do wish to identify schools which have maintained fairly stable educational goals over the years, particularly as these relate to preparation for college. Please characterize your school using the scales below.

In general, the meaning of academic excellence, such as a superior rank in class, was

☐	☐	☐	☐
Very Unstable	Unstable	Stable	Very Stable

The importance attaching to academic excellence, in the sense of recognition for it, was

☐	☐	☐	☐
Very Unstable	Unstable	Stable	Very Stable

	Yes	No
Has the reading load of your students changed markedly over the last 15 years?	☐	☐
Has the writing load of your students changed markedly over the last 15 years?	☐	☐

In the space below we would appreciate a brief comment on the constancy of standards of academic excellence in your school over the years in question. For example, in computing rank in class are certain kinds of courses excluded? Do you feel that rank in class, academic honors generally have been treated comparably over the years?

____
____
____
____
____
____
____
____
____
____

INSTRUCTIONS FOR SCORE REPORT, FORM C-1

Please complete the attached form, entering the SAT data for the top 10% of the graduating class in each of the eight designated years. Your graduating class will vary in size from year to year, so that a different number of records will be reported for May, 1960, than for 1964. The number of reports should be determined by computing 10% of the class and rounding to a whole number.

The forms provide enough space for a graduating class of 250 students, 25 ranks. However, a few schools may need more entries. Simply attach a second sheet for the year in question.

If the tests were never taken, enter N.A. for "not available". If the tests were taken, but the results cannot be supplied, enter "Missing".

INSTRUCTIONS FOR SCORE REPORT, FORM C-2

Please complete the attached form, entering the SAT data for the top 20% of the graduating class in each of the eight designated years. Your graduating class will vary in size from year to year, so that a different number of records will be reported for May, 1960, than for 1964. The number of reports should be determined by computing 20% of the class and rounding to a whole number.

Each form provides enough space for 25 ranks, simply add additional pages as needed for larger classes.

If the tests were never taken, enter N.A. for "not available". If the tests were taken, but the results cannot be supplied, enter "Missing".

INSTRUCTIONS FOR SCORE REPORT, Form C-3

Please complete the attached form, entering the SAT data for the top 30% of the graduating class in each of the eight designated years. Your graduating class will vary in size from year to year, so that a different number of records will be reported for May, 1960, than for 1964. The number of reports should be determined by computing 30% of the class and rounding to a whole number.

Each form provides enough space for 25 ranks, simply add additional pages as needed for larger classes.

If the tests were never taken, enter N.A. for "not available". If the tests were taken, but the results cannot be supplied, enter "Missing".

FORM C

Score Report

School ____

	Year ____			Year ____		
Rank	V	M	Sex	V	M	Sex
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____
____	______	______	____	______	______	____

APPENDIX B

Letter sent to Schools, September 1976
Outline of Study
Instructions for Score Report, Form C
Form C. Score Report for Valedictorians/Salutatorians

EDUCATIONAL TESTING SERVICE PRINCETON, N. J. 08540

Area Code 609
921 - 9000
CABLE-EDUCTESTSVC
•
Developmental Research Division

September 23, 1976

In late May we sent you a request for some information in connection with a study of declining scholastic aptitude and achievement test scores. As you know, this phenomenon has received much attention in recent years. Concerned by the situation, the College Entrance Examination Board has established a panel of prominent educators to review the factors associated with the decline. Under the supervision of this panel, a number of studies are being conducted.

Our request invites your participation in a study of declining scores for selected populations. As the accompanying material indicates, we are interested in the score patterns of students in schools where there has been little change in the demographic make-up of enrollments. Your school was chosen because census data indicates your community may be considered "stable" in terms of its social and economic characteristics.

In response to communications from the schools and the panel of educators, we have modified our request somewhat since May. We now seek information about sixteen students: the valedictorian and salutatorian in each of the even-numbered years from 1960 through 1974. We wish to learn the SAT-V and SAT-M scores for these students, or at least their names, so that we could look up their scores in our own records. We have developed three forms for recording information about the characteristics of your community, your school, and the SAT scores. We hope that you will be able to answer the questions about community and school characteristics and that an appropriate test officer can complete the score report forms.

We do not require the names of any individual students, only verbal and math scores, rank in class, and sex. We will not reveal the scores of any school to others. However, we will prepare a report for you, providing your school's score averages, those averages achieved by all schools in the sample, and a list of participating schools.

Your cooperation will help us better understand the score decline. If you have any questions about the study, please call either of us collect at (609) 921-9000.

Sincerely,

Thomas F. Donlon

Thomas F. Donlon
Senior Research Psychologist

Echternacht

Gary J. Echternacht
Research Psychologist

Enclosures

Outline of the Study

Interpreting SAT mean data over a long period of time is made difficult by the shifts in the composition of the group from year to year. As more and more students elect to seek a college education, the groups become more and more diverse. It is possible that a substantial portion of the decline is attributable to this shift in the nature of the group.

No research study can fully solve this problem. However, we seek to reduce variability by focussing on students with specific characteristics in terms of academic attainment. Essentially, this focus is achieved by limiting the sample to students with a defined rank in class. Is the decline observed, for example, in samples of students all of whom were valedictorian students standing first in their class? For salutatorians? If such samples can be identified, it may be assumed that they are less subject to year to year variability than the self-selected groups who attempt College Board Tests. Further, the study attempts to identify schools for which the meaning of rank in class may have been more consistent over the years: schools in communities which have not experienced marked shifts in racial or ethnic composition, in socioeconomic status, in the typical occupations of family breadwinners, etc.

We ask that school supply the scores or names of sixteen students: the valedictorian and salutatorian in each of eight even-numbered years from 1960 to 1974. Our project will combine the reports of individual schools into a single coherent report. The initial phase of our study will focus on approximately 60-75 schools. If it is successful, a larger extension of the study may be proposed.

COMMUNITY CHARACTERISTICS FORM A

We are especially interested in the degree to which the socioeconomic character of your school district and community have changed over the past 15 years. Although few areas of our country have been immune to social and economic change during the past 15 years, some communities and school districts have more or less retained their social and economic characteristics. In particular their school enrollments have neither increased or decreased dramatically, the racial composition of the student body has remained the same, the socioeconomic status of the students has remained the same, and the general economic character of the community has remained the same (e.g., predominanatly agricultural products, manufacturing, mining, retail or wholesale trade, transportation, etc.) In the space provided below, please indicate the degree that you would say your school district has remained socially and economically stable over the past 15 years.

Very Unstable	Unstable	Stable	Very Stable
☐	☐	☐	☐

Also, answer the following specific questions.

	Yes	No
1. Has your school district's enrollment increased or decreased by more than 10% over the last 15 years?	☐	☐
2. Has the ethnic or racial composition of your enrollment changed by more than 3% over the last 15 years.	☐	☐
3. Has the socioeconomic status of your enrollment changed substantially over the last 15 years?	☐	☐

4. Please indicate by checking one box below the major economic activity of your community.

Agricultural Products	Manufacturing	Mining	Retail/Wholesale Trade	Transportation
☐	☐	☐	☐	☐

Other - specify ____

____

____

In the space provided below would you please provide a brief description of the social and economic character of your community and how it may have changed over the past 15 years.

EDUCATIONAL CHARACTERISTICS FORM B

We are interested in changes in the schools themselves, particularly as these center on concepts of academic excellence. It is inconceivable that education has not changed extensively in process during the years 1960-75. Our study cannot assume therefore, that the best students of 1960 are equivalent to those of 1974. But we do wish to identify schools which have maintained fairly stable educational goals over the years, particularly as these relate to preparation for college. Please characterize your school using the scales below.

In general, the meaning of academic excellence, such as a superior rank in class, was

☐	☐	☐	☐
Very Unstable	Unstable	Stable	Very Stable

The importance attaching to academic excellence, in the sense of recognition for it, was

☐	☐	☐	☐
Very Unstable	Unstable	Stable	Very Stable

	Yes	No
Has the reading load of your students changed markedly over the last 15 years?	☐	☐
Has the writing load of your students changed markedly over the last 15 years?	☐	☐

In the space below we would appreciate a brief comment on the constancy of standards of academic excellence in your school over the years in question. For example, in computing rank in class are certain kinds of courses excluded? Do you feel that rank in class, academic honors generally have been treated comparably over the years?

INSTRUCTIONS FOR SCORE REPORT, FORM C

Please complete the attached form, entering the SAT data for the valedictorian and salutatorian in each of the eight designated years.

If the tests were never taken, enter "Not Tested", but indicate the sex of the student. If the tests were taken, but the results cannot be supplied, enter "Missing". If the school had no valedictorian or salutatorian in a given year, enter "N.A." for not applicable.

FORM C

Score Report

School ____

		V	M	Sex
1960	Valedictorian	______	______	______
	Salutatorian	______	______	______
1962	Valedictorian	______	______	______
	Salutatorian	______	______	______
1964	Valedictorian	______	______	______
	Salutatorian	______	______	______
1966	Valedictorian	______	______	______
	Salutatorian	______	______	______
1968	Valedictorian	______	______	______
	Salutatorian	______	______	______
1970	Valedictorian	______	______	______
	Salutatorian	______	______	______
1972	Valedictorian	______	______	______
	Salutatorian	______	______	______
1974	Valedictorian	______	______	______
	Salutatorian	______	______	______

A Comparative Study of Secondary Schools with Different Score Patterns

Gary J. Echternacht
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

January 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $3 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

ABSTRACT

In this study curriculum, institutional, teacher, and student factors associated with those schools having the largest decrease in Scholastic Aptitude Test (SAT) score averages were compared with those factors associated with schools having increasing or steady SAT score averages. It was believed that by identifying these factors, especially those suggested by a listing of hypotheses for the decline developed by the Education and Society Subcommittee of the Advisory Panel on Score Decline, some insight into the role that schools have played in the score decline might come to light. Although some significant differences between the two groups were found in terms of enrollments in academic courses, ability grouping, age of schools, dropout rate, and teacher experience, these were not believed to be an explanation of a large portion of the decline. Indeed the role of schools in the score decline appeared small.

BACKGROUND OF THE STUDY

The idea that in some way high schools are partly to blame for the decline in achievement test scores is often heard. Many who criticize current practices in education are quick to cite the recent decline in Scholastic Aptitude Test (SAT) score averages in support of their claims. In its effort to explain the steady drop in SAT-verbal and SAT-mathematical scores that has occurred each year since 1963, the Advisory Panel on Score Decline has looked in four areas for explanations of the decline. Change in the content of the SAT or errors in the processes used to equate scores on one form to scores on another, change in the testing population, change in the schools, and change in society have been suggested as possible sources of explanation for the score decline. One of the advisory panel subcommittees, the Education and Society Committee, has been concerned with the extent to which changes in secondary school curriculums, institutional policies, teachers, and students may have resulted in the SAT score decline. This study addressed that concern. Its purpose was to find characteristics that distinguished high schools having a severely declining score trend from those having either a steady or increasing score trend. Such characteristics might suggest some conditions in the schools that have led to the score decline.

What might lead one to believe that school characteristics might be a source of the score decline? Frank Armbruster in a draft copy of his report on elementary and secondary education has argued that school policies should be blamed for this decline in scores.[1] Armbruster notes that scores have declined on all types of achievement tests, not only the SAT but also the various achievement tests used in state assessment programs.[2] The resulting patterns of decline have been remarkably similar. Using test data beginning in the 1930s, Armbruster systematically rejects socioeconomic factors, changing family attitudes, and changing home environments as adequate explanations of the decline. He then asserts that "[t]he 'national IQ' . . . seems to have held relatively constant until 1970." This evidence, he concludes, tends to undermine the theory that "an inherently less talented cohort [is] moving through the school system." Having eliminated all possible sources of the decline other than the school, Armbruster concludes that the source of the decline must lie within the school.

PROCEDURES

Two groups of high schools were identified for the study. One group had undergone a severe decline in SAT scores during the past 10 years; another group had undergone either a steady or even an increasing trend in SAT scores during the same time. Responses to a questionnaire concerning changes in curriculum, institutional policies, students, and teachers were compared for the two groups. Presumably, some clue as to which school factors might be responsible for the score decline would be gained from those questionnaire items where responses differed significantly for the two groups.

Schools in the study were selected by examining SAT score averages earned by students in high schools in 1965, 1969, 1973, 1974, 1975, and 1976.[3] The score data were contained in two files. There was a special data file that contained both SAT score averages and summary statistics from the Student Descriptive Questionnaire (SDQ) for schools having at least 100 SAT registrations in 1973 and 1974. The SDQ asks students to report on their high school curriculums, grades, and interests and is not required when registering for the SAT. The combined SAT and SDQ file contained score averages for each school from 1973 to 1976. Prior to 1973, score data were contained only on yearly history files, which listed examinees and their scores alphabetically.

The only feasible method for gathering a sample of schools was to select a large pool of schools from the 1973-1976 file and then obtain score averages for those schools from the 1965 and 1969 history file. To have used any other method, such as first sorting all students in a file by school for each year, would have been too costly. After completing this process, schools included in the study were further selected on the basis of their score averages for 1965, 1969, 1973, 1974, 1975, and 1976.

Approximately 3,000 high schools were contained in the 1973-1976 file. Only four of those schools were found to have a consistent score gain trend from 1973 to 1976. The remaining schools had a score decline in either verbal or mathematical in at least one year. Therefore, the steady or increasing score average group was selected on the basis of the number of times a school gained on either the verbal or mathematical score average. A total of 331 schools had score increases in at least four of the possible six times. Those schools formed the steady or increasing pool.

Schools with decreasing score averages were easy to find. Two hundred and forty-three schools were found having declines in three consecutive years of at least one-half of a standard error on both the verbal and mathematical scores. The 99 schools having the largest declines made up the pool of schools with declining scores.

The final sample of schools was selected from the pools by a visual inspection of the score trends. Sixty schools were selected from each group.[4] Ten of the schools in the declining score average group were selected because their scores had the most severe decline. The remaining schools were selected randomly from the pool.

The selection of the steady score schools involved a more complex process. When the 1965 and 1969 scores were combined with those from 1973-1976 for this group, trends from 1965 to 1973 were different from those trends in the 1973-1976 data. Some schools had experienced a decline from 1965 to 1973 and an increase from 1973 to 1976. Most schools had lower SAT scores in 1973 than in 1969. No school in the pool had a consistent, increasing score trend from 1965 through 1976.

In light of the difficulties in defining consistent score patterns for this group, the following selection process was used. The four schools with consistent score increases between 1973 and 1976 were included. The remaining 56 schools were selected because their 1965 and 1969 scores were close to their 1973 scores. Twenty-six of these schools had increases in five out of six times in 1973-1976, while 30 had increases in only four of six times.

The principal of each selected school was mailed a questionnaire. More than 90 percent of the principals who returned questionnaires said that they had been at the same school in 1965. If the principal had come to the school after 1965, he was asked to consult with either a guidance counselor or teacher who had been there in 1965. The questionnaire was developed from a list of hypotheses drafted by the Education and Society Subcommittee in February 1975, which identified possible causes of the SAT score decline related to school and society.[5] Among the hypothesized causes of the decline related to changes in schools were:

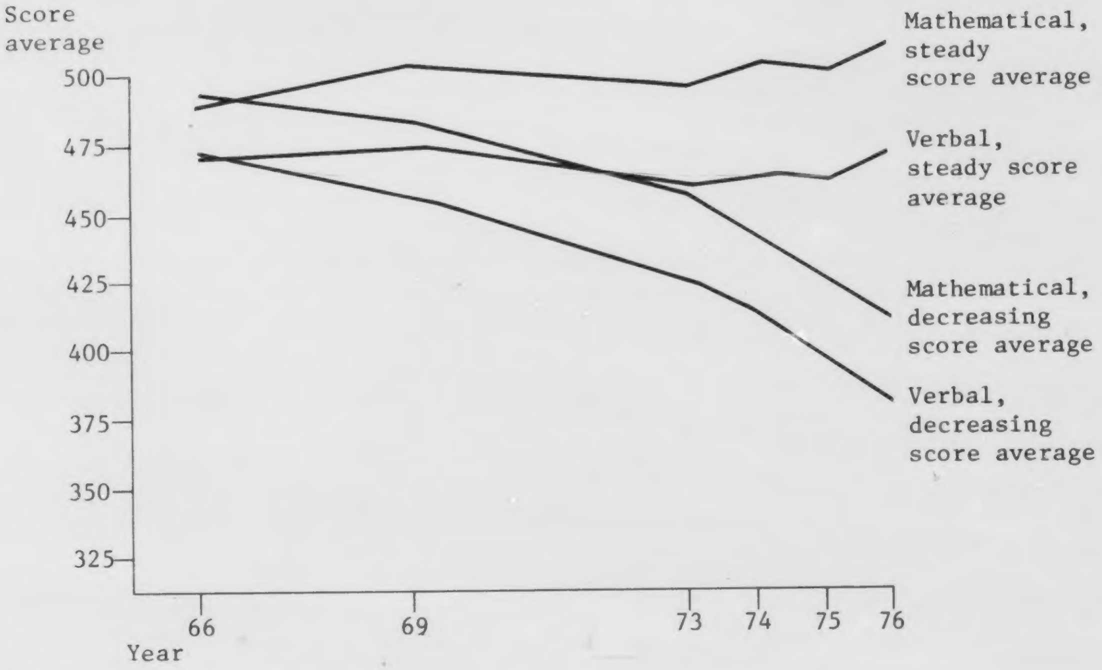
1. Curriculum changes--e.g., decline in foreign language study, decreased emphasis on reading and writing skills, and open classrooms.
2. Institutional policies--e.g., de-emphasis on grades, less homework, increased retention in high school, and failure to "track" students.
3. Teachers--e.g., more clerical aides, less contact with students, and permissiveness of teachers.
4. Students--e.g., decline in competitiveness, truancy, alcohol, and drug abuse problems.

An attempt was made to write a questionnaire item concerning each hypothesis. Sometimes this was impossible. Principals either were thought not able to provide reliable answers to some questions or the hypotheses concerned areas not directly associated with the high school (e.g., whether reading programs used phonics). A copy of the questionnaire is appended.

Characteristics of the Schools and Examinees

Because the SAT history file contained no background data for examinees in 1965 and 1969 during the process of selecting schools, it was impossible to control for pre-existing differences between the schools with steady and decreasing score averages in terms of sex, racial com-

Figure 1. Score Averages for Schools in the Sample



position, and socioeconomic status. Figure 1 shows the score trend in SAT-verbal and mathematical averages for the schools returning questionnaires.[6] The figure shows that each group had about the same verbal and mathematical scores in 1965. From that point on, the score average for the decreasing group of schools began to drop. This drop has continued through 1976, for a total of almost 100 points, or one standard deviation, in 10 years. Meanwhile, the steady score average schools maintained, or slightly increased, their scores during the same time.

How did the two groups of schools differ? Analysis of questionnaire responses indicated that the schools with steady scores currently have slightly higher socioeconomic status than schools with decreasing scores, although the difference was not statistically significant.[7] According to principals in the decreasing score schools, an average of about 20 percent of their school's enrollment in 1965 consisted of minority students.[8] This average had grown to be about 35 percent in 1976. Large minority enrollments were not found in the schools with steady score averages. About 6 percent of their 1965 enrollment and 8 percent of their 1976 enrollment were minority students. Not only did minority enrollment in the decreasing score schools increase between 1965 and 1976, but the proportion of students who were minority students and took the SAT from 1973 to 1976 increased as well. Analysis of the SDQ data shows that in 1973, 26 percent of the SAT examinees from the schools with decreasing scores were minority students. This grew to 37 percent in 1976. These statistics should be compared with 5 percent SAT examinees for both 1973 and 1976 in the steady score average schools.

The academic backgrounds of students who have taken the SAT may have changed also. In 1973 about 21 percent of the students from each type of school said they were ranked in the top 10 percent of their class according to the SDQ responses. By 1976 only 16 percent of the students from schools with decreasing scores said they were in the top 10 percent of their class. The percentage of students in schools with steady scores who said they were in the top 10 percent grew from 21 to 23 percent. The difference between the two groups in 1976 and the difference between the two changes from 1973 to 1976 were both statistically significant.

Grade-point averages as reported on the SDQ were also different for the two groups.[9] Students taking the SAT in schools with steady scores tended to have higher grade-point averages than did those taking the SAT in schools with decreasing scores. This was true for both years. Nonetheless, while grade averages were about the same in the two years for schools with decreasing scores, grade averages for the schools with steady scores had increased.

Changes in Curriculum

Although there is little existing empirical evidence to attribute the decline in test scores to changes in the school curriculum, change in the school curriculum is often cited as a cause of the score decline.[10] In a report on the achievement test score decline, Harnischfeger and Wiley reported statistics from a National Center for Educational Statistics curriculum survey that indicated there have been declining enrollments in traditional academic courses, such as advanced mathematics, English, and foreign language courses.[11] Armbruster places more emphasis for the decline on such things as the attitudes of educators, institutional policies, and teachers. Nevertheless, he concludes ". . . considerable changes will probably be necessary in our current curriculum, particularly if we must reinstitute teaching methods which have been successful in the past."[12]

The members of the Education and Society Subcommittee identified several possible hypotheses from research reports and informal communication regarding how curriculum factors may have caused the score decline. Possible causes of the decline that were identified included the neglect of basic skills, effects of "new math," decline in academic course enrollments, use of newer textbooks that are more concerned with style than with content, and the like.

Analysis of the questionnaire items bearing on the curriculum hypotheses leads to the conclusion that changes in the curriculum explain little of the SAT score decline for this study's sample of schools. Both the schools whose scores have decreased and those whose scores have increased have adopted audiovisual aids, new textbooks, open classrooms, and nontraditional course offerings--e.g., modern film, photography, and studio arts--to about the same degree.[13] Both have been slow in adopting the emphasis of concepts over calculations in mathematics classes.[14]

Only 24 percent of the schools with decreasing scores and 18 percent of the schools with steady or increasing scores indicated that they made a major effort to emphasize concepts rather than calculation in mathematics classes.

Adoption of work-study programs and pass-fail graded courses did not account for any significant difference between the two groups of schools. In fact schools with steady score averages were more likely to have made a major effort to adopt work-study programs and pass-fail graded courses than were schools with decreasing score averages. Seventy-seven percent of the steady score average schools had made a major effort to adopt work-study programs, compared with 42 percent for the decreasing score schools. These statistics fail to support the committee's hypothesis that independent study--i.e., work-study programs and pass-fail graded courses--lead to a decline in score averages.[15]

A statistically significant result was found for the questionnaire item regarding "basis skills." The result, however, was not what was expected from the hypothesis. Rather than an indication of neglect of the "three Rs" by schools with decreasing score averages, just the opposite was found. Sixty-one percent of schools with decreasing scores said they had made a major effort to emphasize teaching the basic skills, while only 37 percent of the schools with steady score averages said the same.[16] This result leads one to speculate that the new emphasis on basic skills represents more a reaction to the score decline than a cause of the score decline.

The increased popularity of the "back to basics" curriculum on the part of schools with decreasing score averages is further illustrated by findings concerning what is taught in English classes. Although the results are not statistically significant, it appears that schools with decreasing scores may be placing more emphasis in their English classes on vocabulary development, grammar, sentence structure, and punctuation.[17] Both decreasing and steady score average schools have emphasized essay writing, reading nineteenth century literature, and reading contemporary literature to the same extent.[18]

Enrollments in foreign language classes have been decreasing throughout the country during the time the SAT score average has declined. Responses from the questionnaire indicate that foreign language enrollments have declined about the same extent in both decreasing and steady score average schools. Only 39 percent of the schools with decreasing scores and 49 percent of the schools with steady scores said their enrollments in modern foreign language--e.g., Spanish and French--have not declined appreciably in the last 10 years. Only 15 percent of the schools with decreasing scores and 24 percent of the schools with steady scores said their enrollments in the classical languages have not appreciably declined. Differences in the enrollment patterns between the two types of schools were not found to be significant.[19]

There was some evidence from mathematics course enrollments that indicated a difference between schools with decreasing and steady score averages. Enrollments in higher mathematics courses had declined to a greater extent for schools with decreasing score averages than they had for schools with steady score averages. For example, 39 percent of the schools with decreasing score averages compared with 18 percent of the schools with steady score averages indicated a decline in enrollments in advanced mathematics courses since 1965. More striking, 33 percent of the schools with decreasing score averages compared with only 6 percent of the schools with steady score averages said their enrollments in geometry had declined in the past 10 years. The finding for geometry, which suggests a possible cause of the score decline, was statistically significant.[20] No significant differences between schools with decreasing and steady score averages were found in enrollment changes for senior year English and nonacademic courses such as journalism and creative writing.[21]

Some evidence that may clarify the role that course selection plays in the score decline was gathered from the analysis of SDQ responses to a question regarding the number of years students took courses in various academic areas. In general, responses to the SDQ indicate that students from schools with steady score averages took about the same number of years of English, mathematics, foreign languages, and physics in 1976 as they did in 1973. Not only did their years of study in these areas remain about the same, but students from steady score average schools also had more years of study in these areas than did students from decreasing score schools. Students from decreasing score schools had more courses in biology during both years, however. Summary statistics are presented in Table 1. Table 2 gives summary statistatics for the questionnaire items regarding curriculum.

Changes In Institutional Policies

Unlike the findings concerning curriculum changes, more statistically significant differences were found between the groups on questionnaire items covering institutional policies. Some of the significant findings, however, contradicated the committee's hypotheses. For example, increasing school enrollment, suggested as a possible cause of the score decline, was more characteristic of those schools with steady score averages than it was of those schools with decreasing score averages. Enrollments in schools with decreasing score averages were about 1,800 in 1965 compared with 1,740 in 1975; enrollments in schools with steady score averages were about 1,230 in 1965 compared with 1,440 in 1975. Although enrollments in schools with decreasing scores were greater than those for schools with steady scores in 1965, a time when both types of schools had about the same scores, enrollments in the former have declined slightly, while enrollments in schools with steady scores have increased about 17 percent.[22] At the same time, both groups of schools have had to reduce their teaching staffs.[23]

No evidence was found to support the hypothesis that increased retention in high school--indicated by a declining rate of dropout from high school--contributes to the SAT score decline.[24] The thought underlying the hypothesis is that as lower-achieving students drop out, instruction can be geared more to the higher-achieving students who typically take the SAT so that higher test scores will result. Analysis of the questionnaire data leads to a different conclusion. Fifty-two percent of the schools with decreasing score averages reported increases in their dropout rate since 1965 compared with that of only 12 percent for steady score average schools.[25] This result suggests that rather than having a positive influence on SAT scores, an increase in the dropout rate indicates either a change in the type of students in the school or some problem within the school.

Table 1. Average Number of Years Courses Were Taken in Six Areas***

Course/year	Male: Decreasing score average	Male: Steady score average	Female: Decreasing score average	Female: Steady score average
English:				
1973	3.9	4.0	3.9	4.0
1976	3.8*	4.0	3.9	4.0*
1973-1976.	.1	.0	.0	.0
Mathematics:				
1973	3.5	3.6	3.2	3.2
1976	3.4	3.6**	3.1	3.3*
1973-1976.	.1	.0**	.1	-.1**
Foreign Language:				
1973	2.5	2.3	2.7	2.7
1976	2.2	2.5	2.4	2.6
1973-1976.	.3	-.2**	.3	.1*
Physical science:				
1973	1.5	1.8**	1.0	1.2*
1976	1.6	1.9**	1.3	1.5**
1973-1976.	-.1	-.1	-.3	-.3
Biological science:				
1973	1.4	1.3**	1.4	1.3*
1976	1.6	1.4**	1.6	1.4*
1973-1976.	-.2	-.1	-.2	-.1
Social science:				
1973	3.3	3.3	3.1	3.3
1976	3.1	3.2	3.0	3.2
1973-1976.	.2	.1	.1	.1

*Indicates significant difference between decreasing and steady score schools, p = .05

**Indicates highly significant, p = .01

***Based on 4 years of secondary school; entries are rounded to one decimal

Analysis of the data regarding institutional policies also leads one to reject hypotheses such as the "edifice complex,"[26] a de-emphasis on grades, and less homework as contributing much to the SAT score decline. Although schools with decreasing score averages were found to have older buildings than those with steady score averages (73 percent of the schools with decreasing scores said their buildings were over 20 years old, while only 37 percent of the schools with steady scores said the same), only two schools from the steady score group had buildings less than five years old, and only 11 had buildings less than 10 years old.

The average per-student cost, which is related to the notion of the "edifice complex," was found to be virtually identical for the two groups in 1976.[27] Nevertheless, if schools with decreasing score averages were getting more lower socioeconomic status students, they may have needed

Table 2. Percentages of Schools in Each Group Responding as Indicated

Item response	Percent decreasing score average	Percent steady score average	Statistically significant difference
Have made a major effort to adopt or expand:			
Basic skills	60.6	36.7	Yes
Audiovisual aids	45.5	49.0	No
Essay writing	45.5	40.8	No
Vocabulary development	48.5	28.6	No
Work-study programs	42.4	67.4	No
New textbooks	60.6	51.0	No
Concepts in mathematics	24.2	18.4	No
Contemporary literature	39.4	16.3	No
Have made some effort to adopt or expand:			
Content in writing	42.4	612.	No
Pass-fail grading	27.3	28.6	No
Open classroom	48.5	40.8	No
Nontraditional courses	72.7	65.3	No
19th century literature	60.6	59.2	No
Schools indicating a decline in enrollment:			
Modern foreign language	60.6	51.9	No
Classical languages	84.9	75.5	No
Advanced math	39.4	18.4	No
Geometry	33.3	6.1	Yes
Senior English	15.2	24.5	No
Journalism, creative writing	27.3	16.3	No

increased funds per student; if they did not receive these funds, that might in part account for their decreasing scores. No evidence was found substantiating the claim that a de-emphasis on grades explains much of the decline. There was general agreement among all the schools in each group that students in their schools in 1965 were more motivated to get higher grades than were students in their schools in 1976. Only six percent of the schools with decreasing scores and 10 percent of the schools with steady scores believed their students were more motivated to get high grades in 1976 than in 1965.[28]

As with the emphasizing of "basic skills," school policies on homework had changed to a different extent for the two groups. Perhaps as a reaction to the score decline, schools with decreasing scores indicated that they had placed more emphasis on homework assignments.[29] Thirty-nine percent of the schools with decreasing score averages compared with only 14 percent of the schools with steady score averages said that they had made a major effort to increase the amount of homework for their students.

One major finding substantiating a hypothesis to the score decline was found regarding how students are grouped within the schools. Thirty-three percent of the schools with decreasing scores said that they had made major efforts to adopt mixed-ability or heterogeneous class grouping during the last 10 years. Ten percent of the schools with steady scores said they had adopted similar efforts. The difference between the two groups was statistically significant.[30]

Results of all the analyses regarding institutional policies are summarized in Table 3.

Changes in Teachers

Most of the principals responding to the questionnaire said that, in many ways, their teachers had changed. Teachers in both groups of schools were reported to have used teacher aides more frequently, been less dedicated, had more formal education, had less desire to enforce high academic standards, been more permissive, spent less time outside

Table 3. Summary Statistics for Questionnaire Items Regarding Institutional Policies

Item response	Decreasing score average	Steady score average	Statistically significant difference
Made a major effort to adopt:			
Mixed ability grouping	33%	10%	Yes
Homework assignments	39%	14%	Yes
1975 students more motivated to get higher grades.	6%	10%	No
Schools over 20 years old	73%	37%	Yes
Increased dropout rate.	52%	12%	Yes
Staff reductions.	45%	45%	No
Cost per student			
1965	$509	$509	No
1976	$1,230	$1,350	No
1976-1965.	$721	$844	No
Enrollment			
1965	1,800	1,230	Yes
1976	1,740	1,440	No
1976-1965	-60	210	Yes

class with students, and placed less emphasis on classroom discipline in 1976 than their counterparts did in 1965. These changes were found equally in schools with decreasing and steady score averages. No statistically significant differences were found between the two groups of schools on items concerning use of teacher aides, dedication to teaching, education beyond the bachelor's level, desire to enforce academic standards, permissiveness, motivation to spend time with students outside class, desire to be liked by students, and emphasis on discipline.[31]

Analysis of the questions that asked principals to compare the teachers in their schools in 1976 with those of 1965 resulted in finding only one statistically significant difference between the two groups of schools.[32] In general, teachers in schools with steady scores tended to have more experience than did teachers in schools with decreasing scores. Forty-seven percent of the schools with steady scores reported that having teachers with five or more years' experience was more common in 1976 than in 1965. Twenty-one percent of the schools with decreasing score averages reported the same. It should be kept in mind that between 1965 and 1976 enrollments at the decreasing score average schools remained steady, or declined slightly, while enrollments at the steady score average schools increased. On the one hand, such a change in enrollments suggests that more new teachers would be added to the steady score schools, resulting in fewer teachers having more than five years' experience in 1976. On the other hand, results concerning teaching experience suggest that teacher turnover is greater in the schools with decreasing score averages than in the schools with steady score averages. A large group of teachers may have retired at the decreasing score average schools, which could account for the result. Another possibility, which is supported by other evidence, is that conditions in the school environment--e.g., a different type of student--may contribute to the relatively higher turnover at the decreasing score schools.

A summary of the findings concerning teachers appears in Table 4.

Changes in Students

Student populations in 1976 have changed since 1965, just as teachers have changed. The majority of principals responding to the questionnaire said that students in their schools in 1976 were more often truant, discipline problems, alcohol and drug abusers, and delinquent than students in their schools were in 1965. Students in 1965 were perceived as competing harder to get into certain colleges, motivated more to score high on college admissions tests, feeling more that education was important preparation for a good job, and enjoying educational activities for their intrinsic values. The perceptions reported by principals from schools with decreasing scores paralleled those reported by principals from steady score schools.[33] Alcohol and drug problems were cited more frequently as being major problems in 1976 for the steady score average schools, although the difference between the two groups of schools was not statistically significant.

Summary statistics are presented for the questionnaire items concerning students in Table 5.

Table 4. Summary Statistics for Questionnaire Items Concerning Changes in Teachers

Questionnaire item	Percent decreasing score average	Percent steady score average	Significant difference
Made a major effort to use teacher aides	12	20	No
More characteristic of teachers in 1976 than in 1965:			
Dedication.	6	8	No
Experience (5 years).	21	47	Yes
Education beyond bachelor's	55	74	No
Enforcing academic standards. . . .	0	8	No
Permissiveness.	70	82	No
Reliance on aides	52	57	No
Spend time with students outside of classroom	21	16	No
Desire to be liked by students. . .	46	45	No
Emphasis on discipline.	6	2	No

Table 5. Summary Statistics for Questionnaire Items Concerning Students

Questionnaire item	Percent decreasing score average	Percent steady score average	Significant difference*
More characteristic of students in 1976 than in 1965:			
Truancy	82	65	No
Discipline problems	64	59	No
Alcohol and drug abuse.	70	88	No
Crime, juvenile delinquency	82	71	No
More characteristic of students in 1965 than in 1976:			
Competition to get into college . .	58	39	No
Motivation to score high on tests .	58	37	No
Education as preparation for job. .	46	41	No
Enjoyment of education.	58	35	No

*Chi-square values calculated in the order they are given in Table 5 are 2.66, .16, 4.09, .59, .01, 3.82, .90, and 4.22.

How Principals Interpret the SAT Score Decline

The SAT means different things to different people. The College Board says that the SAT provides a measure of verbal and mathematical reasoning abilities that are needed to get good grades in college. A student's score provides a valid indication of how well he will do in college. The SAT means other things to other people. Many say that the SAT, or for that matter any standardized test, reflects the culture and family background of the student more than his ability to do college work.[34] The SAT measures a general level of academic achievement attained during high school according to those who use the SAT score decline as evidence of an invasion of rash or foolish new teaching methods or school policies in the school.

This difference of opinion--on what SAT scores indicate--is reflected in the responses to the questionnaire. When asked to indicate which of three characteristics--general achievement in high school, ability to do college work, and cultural or family background--was best measured by the SAT, 27 percent of the entire sample said the SAT best reflects the cultural and family background of the student. This feeling was most prevalent among principals of schools with decreasing scores, where 41 percent so indicated, compared with only 17 percent for the steady score schools. Fifty-five percent of the principals from steady score average schools said that they believed that the SAT best reflected a student's "ability to do college-level work." Only 28 percent of the principals from the declining score average schools said the same.[35] One can see that the principals from the two groups of schools believe the SAT measures different things. Principals from schools with decreasing scores believe that the SAT best reflects the cultural and family background of the students; principals from schools with steady scores believe that the SAT best reflects the ability to do college work.

The issue of the SAT score decline and what to do about it have received differing amounts of attention in the sample of schools. The score decline has naturally been a more important issue in the schools with decreasing scores than it has in the schools with steady scores. For example, the SAT score decline was reported to be a major issue in 85 percent of the schools with decreasing scores. Only 18 percent of the steady score average schools said that the score decline was a major issue. Although the SAT score decline has been an issue in the decreasing score school districts, most principals believe that it will have only a moderate impact on educational practices in their districts. Of those principals who said that the SAT score decline was a major issue, only 13 percent said that they believed the decline would have a major effect on educational practice. Apparently principals in schools with declining scores feel that the factors causing the score decline are not under their control.

Summary and Comments

In this study few of the hypotheses advanced for the score decline by the Education and Society subcommittte were supported by the results.

Curriculum factors seemed to exert only a moderate effect. Although both schools with declining score averages and those with steady score averages reported similar decreases in academic course enrollments (with the exception of geometry enrollments), there was some evidence from the analysis of the SDQ that SAT examinees from schools with steady scores now take more academic courses (e.g., English, mathematics, foreign language, and physical science) than SAT examinees in decreasing score schools. The trend has been for examinees in steady score schools to take more of these courses than their predecessors, while examinees in decreasing score schools either take fewer or about the same number of these courses than their predecessors. Open classrooms, independent study, and new methods for presenting material seem to have had little effect on the score decline.

Institutional policies for the two groups of schools provided for the best discrimination between the two groups. Schools with steady score averages tended to use homogeneous grouping more frequently, had newer school buildings, and have a decreasing dropout rate. Schools with decreasing score averages tended to use heterogeneous grouping more frequently, have older school buildings, and an increasing dropout rate. In an apparent attempt to change the declining achievement trend in their schools, homework assignments and basic skills have been emphasized by schools with decreasing score averages.

Few significant differences were found between teachers in the two groups of schools. Teachers in both groups were generally regarded as less dedicated, more permissive, and less willing to enforce academic standards in 1976 than they were in 1965. In 1976, however, schools with steady score averages reported having had more teachers with five or more years' experience than in 1965. This increase in experience was a significant characteristic of schools with steady scores.

Perceptions of how students have changed since 1965 were also similar for both groups. In 1976 students were viewed as being less motivated to score well on tests and having more behavioral problems (e.g., alcohol, truancy, juvenile delinquency). Nonetheless, there was no significant difference between the two groups.

What might be made of these results? One might say that little has been proved. The questionnaire that was used may have been insensitive to many school effects causing the decline. The sample size may have been too small. The principals might have given unreliable or invalid responses. Each of these is possible.

To the author's knowledge, only one other study has been performed that uses the methodology employed in this study.[36] In that study a sample of Massachusetts schools was chosen and the investigation was limited to four variables: (1) number of academic courses offered; (2) enrollments in science fiction and radio and television courses; (3) age, years of experience, and educational level of the teaching staff; and (4) pupil-teacher ratios. The results of that study indicated that schools with higher rates of decline had fewer course offerings in the academic subjects, presumably indicating higher enrollments in these courses. If this were so, it would support the findings of this study, at least in the case of mathematics and geometry course enrollments. No significant differences were found for the course enrollments, teacher variables, or pupil-teacher ratios in the Massachusetts study. With the exception of the finding regarding teaching experience, the results of this study are much the same as were those of the Massachusetts study.

Some significant factors associated with schools that may explain the SAT score decline were found in this study. Nevertheless, these factors act in complex ways that cannot as yet be explained. Some factors may be causes of the decline, while others are effects of the decline. Taken by themselves, they cannot explain the score decline. Although one might quibble over the degree of assuredness in the remarks by Sidney P. Marland, Jr., president of the College Board, on the effects of schools on the score decline that he made in announcing the creation of the Advisory Panel on Score Decline, it is difficult to deny his assertion that "To single out the schools as being responsible for the decline is unwarranted, unfair, and scientifically unfounded."[37]

NOTES

1. Armbruster, F., The U.S. Primary and Secondary Educational Process. Final draft. HI-2308 RR. Croton-on-Hudson, N.Y.: Hudson Institute, 1975.

2. For example, in both Hawaii and Iowa, score averages have declined on state assessment tests. Hawaii uses the School and College Ability Test (SCAT) and the Sequential Test of Educational Progress (STEP) as its assessment instrument; Iowa uses the Iowa Test of Educational Development (ITED) in its assessment.

3. The years designated refer to school years ending in the designated year. Thus, 1965 actually refers to the 1964-65 school year. Because the process used to get score averages for schools from 1963 to 1972 was so expensive, only two years were included from that period.

4. The final selection of 60 schools from each pool was made because it was desired to have the two groups be as different as possible in terms of score patterns. Had more schools been included in the study, the disparity between the two groups would have been reduced, and it was believed that the questionnaire used in the study would not have been sensitive enough to pick up differences between the two groups.

5. Wharton, Yvonne L., Hypotheses Advanced to Explain the SAT Score Decline. Prepared for the Advisory Panel on Score Decline. N.Y.: College Entrance Examination Board, 1977.

6. Eighty-two percent of the schools with steady score averages returned questionnaires. Fifty-five percent of the schools with decreasing score averages returned questionnaires.

7. Socioeconomic status was based on the reports of principals of the amount of ESEA Title I funds received, type of housing found in the community, and the proportion of students from homes earning more than $20,000 a year.

8. Blacks, Mexican Americans, Puerto Ricans, American Indians, and Orientals were considered minority students.

9. Means and significance tests for grade point averages were:

	Steady score	Declining score	t-value	P
a. Average in 1973	3.08	2.98	2.28	.025
b. Average in 1976	3.18	2.98	5.08	.001
c. 1973-1976	-.10	.00	3.54	.001

10. "Crisis in the schools." U.S. News and World Report, Vol. 79, September 1, 1975, pp. 42-48

11. Harnischfeger, A., and Wiley, D., "Achievement Test Scores Drop. So What?" Educational Researcher, March 1976

12. Armbruster, F., The U.S. Primary and Secondary Educational Process, Final draft. HI-2308 RR. Croton-on-Hudson, N.Y.: Hudson Institute, 1975. p. 293. Armbruster says that there was not an agreement among those providing him with advice on how much change in the curriculum would be required to reverse the trend in score averages.

13. This conclusion is based on chi-square analyses of the 2 X R contingency tables having type of school--i.e., declining score average and steady score average--and response category as the classifying variables. In many instances categories of response were combined so that there were at least five responses in each cell of the table. The specific items and their calculated chi-squares are:

1. Use of audiovisual aids in the classroom, $\chi_1^2 = .10$
2. Use of new edition textbooks $\chi_1^2 = .73$
3. Use of open classrooms, $\chi_1^2 = .47$
4. Nontraditional course offerings, $\chi_1^2 = .50$

The subscripts represent the number of degrees of freedom for the test.

14. Questionnaire item "Emphasis on concepts rather than calculation in mathematics classes," $\chi_1^2 = .41$

15. Although the finding regarding work-study programs was strong, it was not statistically significant. Items were:

work-study programs, $\chi_2^2 = 5.13$

pass-fail or credit-only grading, $\chi_2^2 = .02$

16. Questionnaire item: "Emphasis on teaching the basic skills," $\chi_1^2 = 4.5$, $p = .05$

17. Items include:

"Emphasis on vocabulary development," $\chi_1^2 = 3.36$

"Emphasis on content rather than form in essay writing--i.e., decreased emphasis on grammar, sentence structure, punctuation," $\chi_1^2 = 2.79$

18. Questionnaire items include:

"Essay writing in English class," $\chi_1^2 = .17$

"Stress on 19th century literature," $\chi_1^2 = .02$

"Stress on contemporary literature," $\chi_1^2 = 2.91$

19. Questionnaire items asked for an indication of a five-point scale of the growth or decline in enrollments since 1965. Chi-square values were:

a. Modern foreign language, $\chi_4^2 = 1.36$

b. Classical languages, $\chi_4^2 = 1.05$

20. Significance tests showed chi-square values of:

a. Advanced mathematics--e.g., advanced algebra, trigonometry, $\chi_2^2 = 4.64$

b. Geometry, $\chi_2^2 = 10.8$, $p = .05$

21. Senior year English, $\chi_2^2 = 1.01$; Journalism, creative writing, $\chi^2 = 2.28$

22. The difference in 1965 between enrollments for the two groups is significant at $p = .01$. The differences in 1975 are not significant.

23. Although no data were collected on teacher-pupil ratios, each school was asked to report the extent of any staff reductions or increases. Many of these data were reported on the basis of impressions and hearsay. This led to a decision to code the staff reduction question as a dichotomous item: a school either had or did not have a staff reduction during 1965-1976. Chi-square analyses were not significant for this item.

24. See Harnischfeger and Wiley.

25. Respondents were asked to indicate whether the dropout rate for their school had generally increased, decreased, or remained the same since 1965.

26. The "edifice complex" refers to the "delusion that better schools require shinier physical plants, grander stadiums, fancier equipment, etc. (see list of hypotheses)." It is difficult to use this hypothesis to explain the difference between the two groups of schools in this study. If new schools represent a delusion of better education, one might expect schools with decreasing score averages to be more deluded--i.e., have newer buildings--than would schools with steady score averages.

27. Average per-student costs were $509 (1965) and $1,230 (1976) for the declining score schools and $506 (1965) and $1,350 (1976) for the steady score schools. Differences between the costs for the two groups in each year, as well as the gain, were not statistically significant.

28. There was no significant difference between the two groups in their responses. $\chi_2^2 = 2.74$

29. Item was "use of homework assignments to supplement classroom instructions," $\chi_1^2 = 6.74$, $p = .01$

30. $\chi_1^2 = 7.76$, $p = .05$

31. Conclusions were based on chi-square analysis of 2 X 3 contingency tables. Items and chi-square values were:

1. Use of teacher aides	$\chi_2^2 = 3.96$
2. Reliance on teacher aides.	$\chi_1^2 = .01$
3. Dedication to teaching	$\chi_2^2 = .10$
4. Education beyond bachelor's.	$\chi_2^2 = 4.10$
5. Enforce academic standards	$\chi_1^2 = 1.27$
6. Permissiveness	$\chi_1^2 = .01$
7. Spend time with students	$\chi_1^2 = 1.09$
8. Desire to be liked by students	$\chi_1^2 = .10$
9. Emphasis on discipline	$\chi_1^2 = .15$

32. $\chi_2^2 = 8.81$, $p = .05$

33. None of the statistical analyses yielded significant values of chi-square.

34. See, for example, E. Sharp, The IQ Cult, New York, Coward, McCann, and Geoghegan, 1972.

35. $\chi^2_2 = 7.74$, $p = .05$

36. Weinman, J. J., Declining Test Scores: A State Study. Draft of a report prepared by the Massachusetts State Department of Education, December 1976, p. 76-81

37. ATP News, January 1976, College Entrance Examination Board, p. 2

School Information Questionnaire

As you may have read, test scores on the two major tests used for college admissions--the SAT and the ACT--have been declining. In the case of the SAT this decline began around 1965 and has continued through the current year. The decline first became apparent in the late 1960's when it was attributed mainly to an increase in the number of college-bound students taking the test. It was believed that more lower scoring students were going on to college and that these students were depressing the score average. In about 1970, however, the number of students taking the SAT began to steady. Nevertheless, SAT score averages have continued their downward trend.

In an attempt to gain some understanding of the score decline, the College Entrance Examination Board established an Advisory Panel on Score Decline chaired by former Secretary of Labor, Willard Wirtz. During the past year, the Panel has sponsored a number of studies aimed at providing information about possible explanations for the decline. Among the possibilities under investigation are: 1) deficiencies in the way the tests are constructed; 2) changes in society; and 3) changes in the school environment. The current study falls into the third area: its purpose is to determine if there is any link between the characteristics, policies, and resources of a school and its prevailing score trend.

On the following pages you will find questions concerning your high school. Please answer each question as accurately as you can. If you need to consult with other school members--e.g., an English teacher, a mathematics teacher, and so on--I urge you to do so. Each of your answers is important to the study. Nonresponse introduces bias into our findings and diminishes the study's ultimate value.

After the data are collected, a report will be developed for the Panel sponsoring the study. You will also receive a copy of this report. At no time will the names of the schools returning questionnaires be revealed; all of your responses will be kept in strictest confidence.

I thank you for your anticipated cooperation in this matter.

Gary J. Echternacht
Research Psychologist

1. What grades does your school serve?

_____a. Grades 9-12

_____b. Grades 10-12

_____c. Grades 11-12

_____d. Other (Specify: ____)

2. About how old is the major classroom building of your high school?

_____a. Less than 2 years old

_____b. 2 to 5 years old

_____c. 6 to 10 years old

_____d. 11 to 20 years old

_____e. over 20 years old

3. Do students in your high school have to meet a certain level of performance on standardized tests before they are allowed to graduate?

_____a. Yes

_____b. No

4. If your school received any Federal funds from the Elementary and Secondary Education Act (ESEA) in the school years 1972-1973 and 1975-1976, please indicate the amount received.

	1972-1973	1975-1976
a. Title I: Education of children from low income families	$__________	$__________
b. Title III: Supplementary education centers and services	$__________	$__________
c. Title VI: Education of the handicapped	$__________	$__________
d. Title VI: Bilingual education programs	$__________	$__________
e. Title VIII: General provisions	$__________	$__________

5. Check the category that best describes the community in which your high school is located:

_____a. Large central city (population over 500,000)

_____b. Large city (population 100,000 to 500,000)

_____c. Medium-sized city (population 50,000 to 100,000)

_____d. Suburb of a central, large, or medium-sized city

_____e. Small city (population 25,000 to 50,000)

_____f. Small town (population 10,000 to 25,000)

_____g. Rural area

6. Check the category that best describes the principal type of housing in your community:

_____a. All single-family units

_____b. Predominantly single-family units; some multi-family units

_____c. Evenly divided between single-family and multi-family units

_____d. Predominantly 2-3 family units

_____e. Predominantly larger (4 or more) multi-family units

7. Please enter the following information for the current school year. (If you are unable to make a reasonably accurate estimate, please leave the item blank.)

_____a. Enter the approximate percentage of students in your high school whose total family income is greater than $20,000/year

_____b. Enter the approximate percentage of students in your high school whose total family income is less than $5,000/year

_____c. Enter the approximate percentage of the community work force that is unemployed

8. Approximately what percentage of the graduating classes of 1970 and 1975 remained in the community to live and/or work in the year immediately following graduation?

_____% a. 1970

_____% b. 1975

9. Using available data whenever possible, please enter the following information for the school years indicated:

   a. Enter the approximate per-pupil expenditure in your school district for the years indicated.

   b. Enter the school tax millage rate for the years indicated.

   c. Enter the approximate percentage of full time teaching staff in your high school who held advanced degrees in the years indicated.

   d. If it was necessary to lay off some of your teaching staff in order to balance the school budget, enter the precentage of teachers laid off in your high school in that year.

   e. If it was necessary to hire additional teachers in your high school (excluding hiring for replacement) please enter the number of teachers hired in each of the years indicated.

	1964-1965	1968-1969	1972-1973	1973-1974	1974-1975	1975-1976
a. Per-pupil expenditures						
b. Tax millage						
c. % Advanced degrees						
d. % Staff reduction						
e. Staff increase						

10. Using existing data whenever possible, please complete the following table. If such data are not available, make a reasonably accurate estimate. If you can make only a "wild guess," please leave the item blank.

a. Enter the total number of students enrolled in your high school during the school years indicated in the column headings.

b. During the same years approximately what percentage of the students were enrolled in an academic or college preparatory program?

c. Approximately what percentage of students were enrolled in a business or commercial program?

d. Approximately what percentage of the students were enrolled in other vocational programs?

e. Approximately what percentage of the students were enrolled in a general program? (Please check to make sure the percentages for rows B-E equal 100%.)

f. Approximately what percentage of your enrollment was made up of minority students--e.g., Black or Afro-American, Mexican or Chicano, Puerto Rican, Oriental, and American Indian?

g. Approximately what percentage of your senior class entered a two-year college in the year immediately following graduation?

h. Approximately what percentage of your senior class entered a four-year college in the year immediately following graduation?

	1964-1965	1968-1969	1972-1973	1973-1974	1974-1975	1975-1976
a. Total school enrollment						
b. Students in academic program (%)						
c. Students in commercial program (%)						
d. Students in other vocational programs (%)						
e. Students in general program (%)						
f. Minority enrollment (%)						
g. Students attending two-year college (%)						
h. Students attending four-year college (%)						

11. Has the percentage of students who drop out from your school—excluding those who transfer to other schools--generally increased or decreased since 1965?

_____a. Increased

_____b. Decreased

_____c. Remained the same

12. The following is a list of curriculum characteristics that your school may have introduced or expanded over the last 10 years. Please read each item and circle the number that most appropriately describes your school's effort in the area.

Circle 1 if your school has made a major effort to adopt or expand the characteristic
Circle 2 if your school has made some effort to adopt or expand the characteristic
Circle 3 if your school has made no effort to adopt or expand the characteristic

1 2 3 Emphasis on teaching the basic skills
1 2 3 Use of audio-visual aids in the classroom
1 2 3 Integration of contemporary issues in social studies classes
1 2 3 Essay writing in English classes
1 2 3 Work-study programs
1 2 3 Emphasis on concepts rather than calculation in mathematics classes
1 2 3 Decreasing the total number of course requirements for graduation
1 2 3 Stress on 19th century literature in English classes
1 2 3 Stress on contemporary literature in English classes
1 2 3 Mixed ability ("heterogeneous") class grouping rather than same-ability ("homogeneous") grouping or tracks
1 2 3 Use of paraprofessionals or teacher's aides
1 2 3 Use of homework assignments to supplement classroom instruction
1 2 3 Use of new edition textbooks
1 2 3 Increasing the total number of course requirements for graduation
1 2 3 Use of "open classroom"—i.e., emphasis on student preferences, nongraded activities, self-help groups
1 2 3 Pass/fail or credit-only grading
1 2 3 Alternative education schools or programs
1 2 3 Nontraditional, nonacademic course offerings—e.g., modern film, photography, studio arts
1 2 3 Decreasing the number of English courses required for graduation
1 2 3 Decreasing the number of mathematics courses required for graduation
1 2 3 Emphasis on vocabulary development in English classes
1 2 3 Emphasis on content rather than form in essay writing—i.e., decreased emphasis on grammar, sentence structure, punctuation

13. Please indicate the course enrollment trends in your high school since 1965 in the following subject areas by circling the appropriate number. (Please consider trends in terms of percentage of the total student body taking the course rather than the actual number of students enrolled from year to year.)

Circle 1 if the enrollment has increased

Circle 2 if the enrollment has remained the same

Circle 3 if the enrollments have decreased by less than 6%

Circle 4 if the enrollments have decreased by 6%-15%

Circle 5 if the enrollments have decreased by more than 15%

1	2	3	4	5	Modern foreign language (French, German, Spanish, Russian)
1	2	3	4	5	Classical languages (Latin, Greek)
1	2	3	4	5	Advanced mathematics (advanced algebra, trigonometry, calculus)
1	2	3	4	5	Geometry
1	2	3	4	5	Senior year English
1	2	3	4	5	Journalism, creative writing

14. Based on your own observations, how would you say that students at your high school in 1976 compare to those of 1965 on the following characteristics? (If you have been at your present school for less than 10 years, please check here ☐ and answer this question in terms of your observations of students in general.)

Circle 1 if this is more characteristic of students in 1965

Circle 2 if there is no difference between 1965 and 1976 students

Circle 3 if this is more characteristic of students in 1976

1	2	3	Strong motivation to get high grades in school
1	2	3	Truancy
1	2	3	Discipline problems
1	2	3	Competition to get into certain colleges
1	2	3	Alcohol and drug problems
1	2	3	School, crime and juvenile delinquency.
1	2	3	Strong motivation to score high on college admission tests
1	2	3	A feeling that education is important preparation for a good job
1	2	3	An enjoyment of educational activities for their own intrinsic value

15. Based on your own observations, how would you say that teachers in your school in 1976 compare to those of 1965 on the following characteristics. (If you were not at this school in 1965 please check here ☐ and respond in terms of your observations of teachers in general.)

Circle 1 if this was more characteristic of teachers in 1965

Circle 2 if there is no difference between the 1965 and 1976 teachers

Circle 3 if this is more characteristic of teachers in 1976

1 2 3 Dedication to teaching

1 2 3 Five or more years of teaching experience

1 2 3 Education beyond the bachelor's level

1 2 3 Desire to enforce high academic standards

1 2 3 Permissiveness in the classroom

1 2 3 Reliance on clerical aides to reduce out-of-class work load

1 2 3 Motivation to spend time with students outside of the classroom

1 2 3 Desire to be liked by students

1 2 3 Emphasis on discipline within the classroom

16. In the following years, which standardized entrance test did most of your college-bound students take?

1964-1965	1970-1971	1975-1976	
____	____	____	a. Most students took the ACT
____	____	____	b. Most students took the SAT
____	____	____	c. Most students took the ACT and the SAT
____	____	____	d. Most students took neither the SAT nor the ACT

17. In the following years, when did most of your college-bound students take the ACT or the SAT?

1964- 1965	1970- 1971	1975- 1976	
____	____	____	a. During their junior year
____	____	____	b. During their senior year
____	____	____	c. About half took the test during their junior year; half during their senior year
____	____	____	d. During both their junior and senior years

18. In your opinion, which one of the following characteristics is most clearly reflected in the SAT or ACT test scores of your students?

_____a. General academic achievement in high school

_____b. Ability to do college level work

_____c. Cultural and family background of the student

_____d. Other (Specify: ______________________________________)

19. To what extent has the decline in ACT and SAT test scores throughout the nation been an issue in your school district?

_____a. It has been a major issue

_____b. It has been discussed, but it has not been a major issue

_____c. It has not been an issue at all

20. Do you believe that the decline in SAT and ACT score averages will have a major effect on practices in elementary and secondary education?

_____a. Yes, it will have a major effect on educational practices

_____b. Yes, it will have some effect on educational practices

_____c. No, it will not affect educational practices

_____d. Have not thought about it

21. In your opinion, what are some of the most important factors explaining the decline in SAT and ACT scores?

____

____

____

____

____

____

____

Summary of Validity Data from the Admissions Testing Program Validity Study Service

Susan F. Ford
Sandy Campos
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

June 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

OVERALL SUMMARIES

Selected data from 11 years of Validity Study Service (VSS) reports, 1964-65 to 1974-75, are summarized in this report. These data include validities of Scholastic Aptitude Test verbal (SAT-V) and mathematical (SAT-M) scores, high school record (HSR), and the three predictors combined; means and standard deviations of both SAT scores; and numbers of students in the VSS analyses. High school record means and standard deviations have not been summarized because of the variety of scales used. Analysis of these data reveals that the SAT scores are predicting as well in the more recent years as in the earlier years; however, HSR validity and multiple correlations have declined somewhat over this period of time.

Sample Selection

The source of the validity data is the VSS data base. (Selected data from all available validity studies have been transcribed from printed VSS reports to the data base.) All college records meeting the following selection criteria are included in the summaries:

1. The number of students was at'least 50;
2. Student groups included were all freshmen, all SAT takers if an all freshman group was not present, or a general liberal arts group if neither of these groups was present;
3. No more than two entering classes had been combined in a record (combined data were assigned to the earlier year);
4. Years of student entry to college were 1964 through 1974;
5. The criterion was freshman grade-point average (FRGPA);
6. The first three predictors, in any order, were SAT-V, SAT-M, and HSR.

Thirteen records were excluded because of data verification problems.

A coeducational college requesting a validity study may elect to have data analyzed for men and women separately and combined; separately, only; or combined, only. Since validities obtained separately and combined may differ, partly because of varying fields of study chosen by males and females, and since a trend in validities for one sex may not occur for the other, validities have been summarized for males, females, and for the combined sexes. Combined data were not available in the data base for all colleges electing separate male and female analyses. All available combined data have been included in this summary. Data from single sex institutions have been included in the appropriate single sex summaries.

Description of the Sample

Information about selected characteristics of the colleges represented in the 1974-75 summaries was obtained from the College Board Marketing Management File. This descriptive data is reported in Table 1 and includes College Board region in which the colleges are located, type of college (two-year or four-year, public or private), and percent minority enrollment. The information on percent minority enrollment may be somewhat out of date for some colleges. However, it provides a general idea of the ethnic composition of the colleges included. Information is given only for those colleges having validity data for 1974 entrants to college, since the Marketing Management File contains primarily 1974 data.

Variables Included

SAT-V, SAT-M, HSR, and all three combined are the predictors for which validity information is presented. The criterion in all cases is freshman grade point average (FRGPA) at the college requesting the validity study. The scale for FRGPA was not necessarily the same from college to college.

As mentioned previously, HSR may be reported on several different scales. Some colleges report class standing and class size. These data are then converted to a 20 to 80 high school rank scale. Others may use percentile ranks or inverted percentile ranks or high school grade-point average, which may be reported on a number of different scales. Starting with classes entering college in 1972, colleges were given the option of using student-reported HSR computed from Student Descriptive Questionnaire (SDQ) responses. Either rank on a 20 to 80 scale or average on a 0 to 4 scale was used in the validity study. Table 2 presents the numbers of colleges electing each of the HSR options (rank or average) in the male, female, and combined summaries, shown later. Generally, high school rank was the more popular measure of HSR, and most colleges preferred transcript to SDQ data.

Descriptive Statistics on the Samples

SAT-V and SAT-M means and standard deviations for the college included in the validity summary have been summarized in Tables 3 and 4, respectively. Means have been grouped into three categories: those lower than 450, those between 450 and 549, and those 550 or higher. There are no clear trends apparent for males for either SAT score. Rather, mean SAT scores at those colleges electing to conduct validity studies over this period of time, seem to have fluctuated inconsistently from year to year. For females and combined sex groups, there is a fairly clear downward trend in SAT-V means, with increasingly larger percentages of the colleges having mean scores in the lowest score category and fewer with mean scores in the highest category. There is no clear trend in SAT-M scores for either the female or combined sex groups.

Standard deviations have been grouped into categories of smaller than 75, 75 to 84, and 85 or larger. There is a tendency for all three samples to be more heterogeneous in ability as measured by SAT-V or SAT-M in the more recent years.

Summary of Validities

Median validities for the 11 years are presented in Table 5 for males, females, and combined sexes. These medians are also plotted in Figures 1, 2, and 3. The male and female summaries include data from single-sex colleges, as well as coeducational institutions; the combined summary is based on data from all colleges for which a combined sex study was done, whether or not a separate sex study was performed. The table includes the four correlations of interest: those for SAT-V, SAT-M, HSR, and the combined predictors.

HSR is the best single predictor in each of the years summarized, and the validity of the SAT-V, SAT-M, and HSR combination is substantially higher than that of HSR alone. Validities are lower for males than for females for all predictors. Within the male and female samples, SAT-M validities are very similar to SAT-V validities, but SAT-V validities are generally higher within the combined sex samples. SAT validities for recent years for the combined samples average about .40 for SAT-V, .35 for SAT-M, .50 for HSR, and .58 for the combined predictors.

No clear trends over time were apparent for SAT validities, although for all groups, SAT-V and SAT-M validities for 1973 and 1974 were among the highest observed. There was a general downward trend in the validity of HSR accompanied by a slight downward trend in multiple correlations.

SUMMARIES FOR SAMPLES GROUPED BY MEAN SAT-VERBAL SCORE AND FOR TWO-YEAR COLLEGES

A further examination of the trends in validity coefficients for SAT-V, SAT-M, HSR, and the three predictors combined was undertaken to compare their respective validities across groups of colleges differing in mean verbal SAT scores. Three subgroups of colleges were defined on the following basis:

1. SAT-V mean less than 450
2. SAT-V mean between 450 and 549
3. SAT-V mean greater than or equal to 550

Table 6 gives the median validity coefficients for male, female, and combined sex samples in these three subgroups for 1964-1974.

An examination of the data contained in Table 6 reveals only minor differences in the median validity coefficients across subgroups of colleges differentiated by mean verbal SAT scores. Although the differences are small, there is a fairly consistent ordering of the groups based on median validities. Colleges with SAT-V mean scores in the middle range tend to have the highest median validities. Those with SAT-V means in the low range tend to have somewhat lower median validities, and the median validities are generally lower still for the highest verbal scoring samples.

One explanation of this consistent ordering of validity coefficients across subgroups of colleges differing in SAT-V mean scores would attribute a portion of the differences in validities to differences in homogeneity of the samples grouped by mean verbal score. It is reasonable to expect the groups with higher scores (presumably those in more selective colleges) to have less variance on the predictors. An increase in subgroup homogeneity on a predictor variable generally results in the lowering of the correlation of that predictor with other variables. Examination of the levels of homogeneity of the three SAT-V subgroups under study reveals the following ordering of subgroups from high to low homogeneity:

	Percentage of colleges with SAT-V standard deviation less than 80 (high homogeneity)
1. SAT-V mean greater than or equal to 550	65 percent
2. SAT-V mean less than 450	44 percent
3. SAT-V mean between 450 and 549	31 percent

This ordering is directly opposite to that found for the validity coefficients and thus lends support to the hypothesis that differences in homogeneity may account (at least in part) for the observed differences in median validities.

In order to assess the validity of the four predictor variables in community colleges, median validity coefficients were computed for the time period 1964-74 for the male, female, and combined sex samples in those community colleges participating in VSS and meeting the study selection criteria described in Part 1 of this paper. It should be noted that the respective number of colleges in the male, female, and combined sex samples for community colleges is relatively small, a factor that should be taken into account in the interpretation of the results. It should also be noted that those community colleges participating in the Validity Study Service cannot be assumed to be representative of community colleges generally. Table 7 gives the median validity coefficients for community college samples and the comparable total sample median validity coefficients. All of the community college median coefficients are lower than the corresponding total sample medians, suggesting that these variables are somewhat less valid predictors of college grade point average in two-year colleges than in four-year colleges. However, despite the consistent differences, this generalization should be considered tentative in view of the cautions about sample size and representativeness expressed above.

Table 1. Number of Colleges Included in 1974 Validity Summaries by College Board Region, Type of College, and Percent Minority Enrollment[1]

Descriptor	Male summary	Female summary	Combined summary
College Board regions:			
New England	17	17	23
Middle States	21	22	30
Southern	15	18	20
Midwestern	10	10	22
Southwestern	3	3	4
Western	4	5	10
Type of college			
Two-year public	0	0	3
Two-year private	2	2	4
Four-year public	32	37	61
Four-year private	36	36	40
Percent minority enrollment			
0 - 4	39	41	59
5 - 9	20	22	28
10 - 19	9	10	16
20 - 29	2	2	4
45	0	0	1
99	0	0	1

1. One college could not be located on the Marketing Management File, one college had invalid data for the type of college, and minority enrollment was unavailable for 19 colleges.

Table 2. Types of High School Record Used by Colleges Included in Validity Summary

Type of high school record	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974
	Male summary										
College rep. average.	8	6	9	14	11	11	35	8	18	12	17
Student rep. average[1]. . .								1	9	5	5
College rep. rank	50	55	59	80	38	47	106	39	49	34	44
Student rep. rank[1]									2	7	4
	Female summary										
College rep. average.	8	10	14	20	12	14	42	11	17	13	20
Student rep. average[1]. . .								1	9	3	4
College rep. rank	49	60	76	81	37	59	112	39	55	42	48
Student rep. rank[1]									3	7	3
	Combined summary										
College rep. average.	12	15	17	11	7	9	27	7	25	26	30
Student rep. average[1]. . .									11	6	12
College rep. rank	50	65	59	51	34	36	102	25	62	52	63
Student rep. rank[1]									3	5	5

1. Option available beginning in 1972.

Table 3. SAT-Verbal and Mathematical Means for Colleges Represented in the Male, Female, and Combined Summaries of Validities

	Percentages of SAT-verbal means that were			Percentages of SAT-mathematical means that were			
Entering year	lower than 450	450 to 549	550 or higher	lower than 450	450 to 549	550 or higher	Number of colleges
Male summary							
1964.	26	62	12	9	59	33	58
1965.	18	69	13	10	62	28	61
1966.	21	57	22	6	47	47	68
1967.	16	66	18	7	48	45	94
1968.	29	55	16	10	51	39	49
1969.	26	66	9	5	66	29	58
1970.	22	65	13	6	55	39	141
1971.	35	54	10	6	58	35	48
1972.	24	60	15	4	56	40	78
1973.	29	62	9	9	50	41	58
1974.	31	53	16	4	56	40	70
Female summary							
1964.	19	54	26	23	61	16	57
1965.	17	56	27	19	61	20	70
1966.	9	62	29	10	68	22	90
1967.	14	67	19	13	70	17	101
1968.	20	63	16	22	59	18	49
1969.	12	71	16	14	71	15	73
1970.	17	64	19	14	68	19	154
1971.	16	63	22	6	73	22	51
1972.	23	62	15	17	63	20	84
1973.	25	65	11	14	74	12	65
1974.	25	57	17	12	65	23	75
Combined summary							
1964.	24	58	18	19	58	23	62
1965.	22	62	15	21	60	19	80
1966.	21	64	14	13	64	22	76
1967.	26	69	5	19	61	19	62
1968.	37	59	5	20	63	17	41
1969.	22	71	7	11	76	13	45
1970.	21	67	12	13	67	19	129
1971.	34	56	9	12	69	19	32
1972.	30	57	13	14	62	24	101
1973.	31	65	3	16	67	17	89
1974.	39	55	6	16	65	18	110

Table 4. SAT-Verbal and Mathematical Standard Deviations for Colleges Represented in the Male, Female, and Combined Summaries of Validities

	Percentages of SAT-verbal standard deviations that were			Percentages of SAT-mathematical standard deviations that were			
Entering year	smaller than 75	75 to 84	85 or larger	smaller than 75	75 to 84	85 or larger	Number of colleges
Male summary							
1964.	26	34	40	22	22	55	58
1965.	31	44	25	16	41	43	61
1966.	32	50	18	29	37	34	68
1967.	36	45	19	34	36	30	94
1968.	27	33	41	29	31	41	49
1969.	22	40	38	16	36	48	58
1970.	23	40	37	20	35	45	141
1971.	21	40	40	21	35	44	48
1972.	15	36	49	17	36	47	78
1973.	21	34	45	19	29	52	58
1974.	13	31	56	10	27	63	70
Female summary							
1964.	26	28	46	12	37	51	57
1965.	29	47	24	24	40	36	70
1966.	41	38	21	40	31	29	90
1967.	30	37	34	30	31	40	101
1968.	31	33	37	33	35	33	49
1969.	30	27	42	33	29	38	73
1970.	21	35	44	27	34	39	154
1971.	24	41	35	31	43	25	51
1972.	14	31	55	21	29	50	84
1973.	17	37	46	20	42	38	65
1974.	7	37	56	11	36	53	75
Combined summary							
1964.	23	44	34	18	29	53	62
1965.	28	40	32	24	24	52	80
1966.	32	42	26	30	33	37	76
1967.	19	42	39	18	26	56	62
1968.	24	34	41	20	29	51	41
1969.	11	36	53	13	29	58	45
1970.	12	31	57	16	29	56	129
1971.	16	41	44	25	31	44	32
1972.	12	25	63	13	30	57	101
1973.	7	36	57	10	27	63	89
1974.	4	31	65	4	24	73	110

Table 5. Median Validities Obtained Each Year for SAT-Verbal and Mathematical Scores, High School Record, and for the Three Predictors Combined[1]

Entering year[2]	No. of coll.	SAT-V	SAT-M	H.S. Rec.	3 Pred. comb.	No. of coll.	SAT-V	SAT-M	H.S. Rec.	3 Pred. Comb.	No. of coll.	SAT-V	SAT-M	H.S. Rec.	3 Pred. comb.
	Validities for male samples					Validities for female samples					Validities for combined samples[3]				
1964.	58	0.33	0.32	0.52	0.56	57	0.44	0.42	0.57	0.65	62	0.37	0.32	0.55	0.62
1965.	61	0.32	0.31	0.48	0.54	70	0.41	0.38	0.57	0.63	80	0.36	0.29	0.56	0.61
1966.	68	0.29	0.28	0.47	0.52	90	0.38	0.34	0.55	0.61	76	0.33	0.28	0.53	0.57
1967.	94	0.30	0.28	0.45	0.51	101	0.40	0.36	0.55	0.62	62	0.36	0.28	0.52	0.57
1968.	49	0.30	0.29	0.48	0.57	49	0.44	0.40	0.54	0.63	41	0.40	0.33	0.56	0.62
1969.	58	0.28	0.30	0.46	0.52	73	0.38	0.35	0.56	0.61	45	0.39	0.33	0.53	0.59
1970.	141	0.29	0.27	0.43	0.49	154	0.39	0.36	0.52	0.58	129	0.37	0.29	0.49	0.56
1971.	48	0.30	0.26	0.44	0.50	51	0.37	0.39	0.47	0.55	32	0.34	0.28	0.48	0.53
1972.	78	0.30	0.32	0.45	0.52	84	0.40	0.41	0.50	0.59	101	0.35	0.33	0.50	0.57
1973.	58	0.35	0.36	0.44	0.54	65	0.40	0.40	0.47	0.55	89	0.38	0.36	0.50	0.56
1974.	70	0.35	0.37	0.44	0.52	75	0.40	0.42	0.47	0.57	110	0.42	0.39	0.50	0.58

1. Data for a total of 13 colleges were excluded because of data verification problems.
2. Refers to fall of freshman year.
3. Combined samples are from colleges that analyzed combined sex data regardless of whether or not separate sex data were analyzed.

Table 6. Median Validity Coefficients for Groups Classified by Verbal Ability and by Sex, 1964-1974

SAT-V mean range	Male	Female	Combined sex
SAT-Verbal validity coefficients			
Less than 450.	.29	.39	.37
450 to 549	.32	.41	.38
550 and greater.	.29	.33	.33
Total sample	.31	.40	.37
SAT-Math validity coefficients			
Less than 450.	.31	.41	.34
450 to 549	.31	.40	.33
550 and greater.	.26	.30	.26
Total sample	.30	.38	.32
High school record validity coefficients			
Less than 450.	.43	.52	.50
450 to 549	.48	.54	.53
550 and greater.	.43	.43	.45
Total sample	.46	.53	.52
Three predictors combined validity coefficients			
Less than 450.	.50	.59	.58
450 to 549	.54	.62	.59
550 and greater.	.48	.53	.52
Total sample	.52	.60	.58
Number of colleges			
Less than 450.	189	152	229
450 to 549	482	545	514
550 and greater.	112	172	84
Total sample	783	869	827

Table 7. Community College and Total Sample
Median Validity Coefficients, 1964-1974

Predictor variables	Male	Female	Combined sex
Community college sample			
SAT-Verbal.	.26	.37	.32
SAT-Math.	.26	.34	.30
High school record.	.32	.47	.48
Three predictors combined	.40	.56	.53
Number of colleges.	31	36	45
Total sample			
SAT-Verbal.	.31	.40	.37
SAT-Math.	.30	.38	.32
High school record.	.46	.53	.52
Three predictors combined	.52	.60	.58
Number of colleges	783	869	827

Figure 1. Male Sample Median Validity Trends 1964-1974

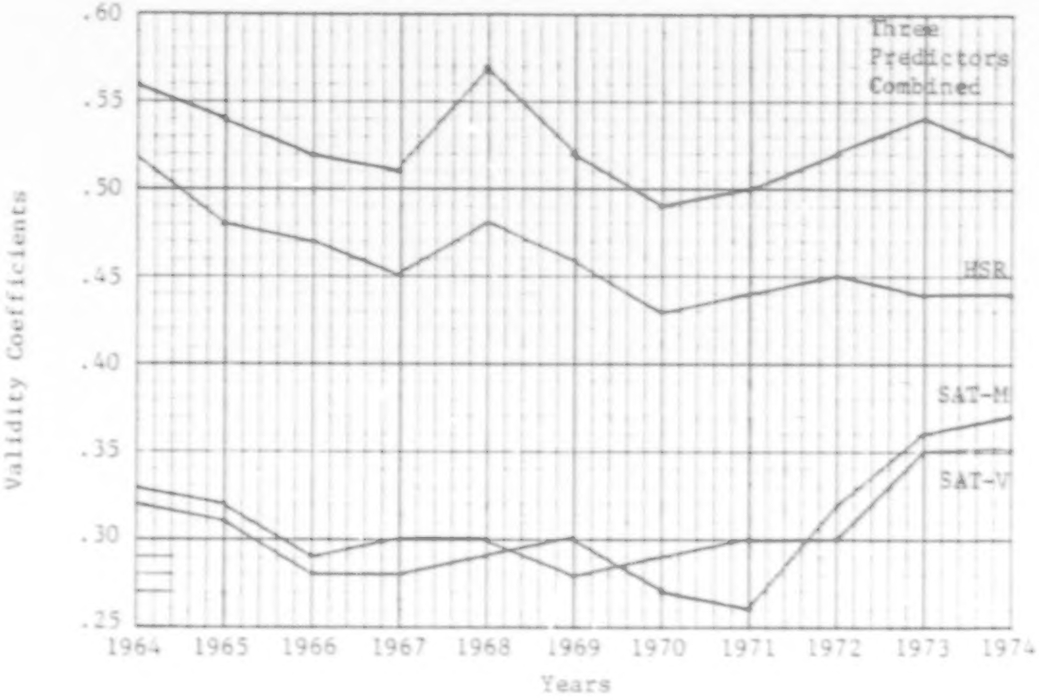


Figure 2. Female Sample Median Validity Trends 1964-1974

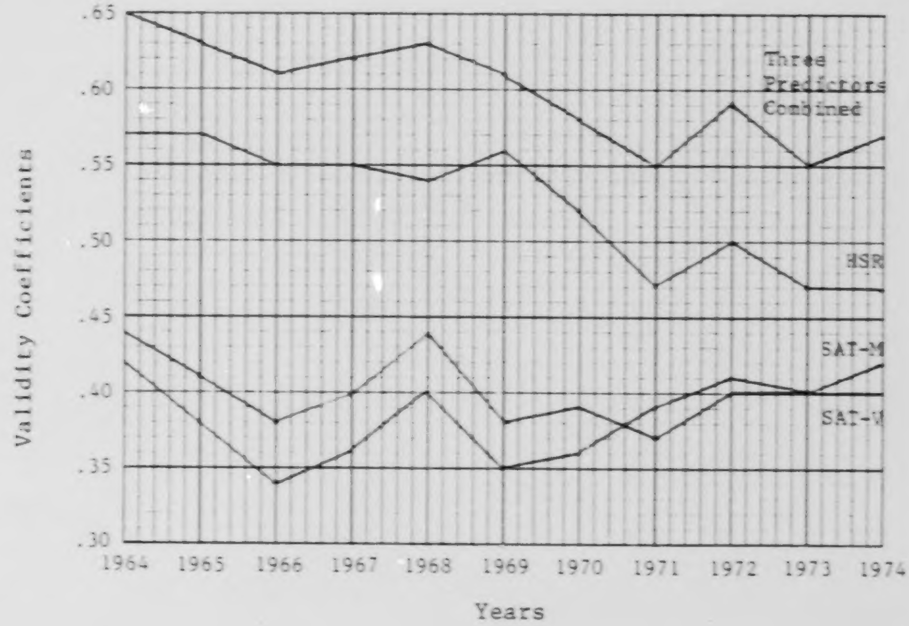
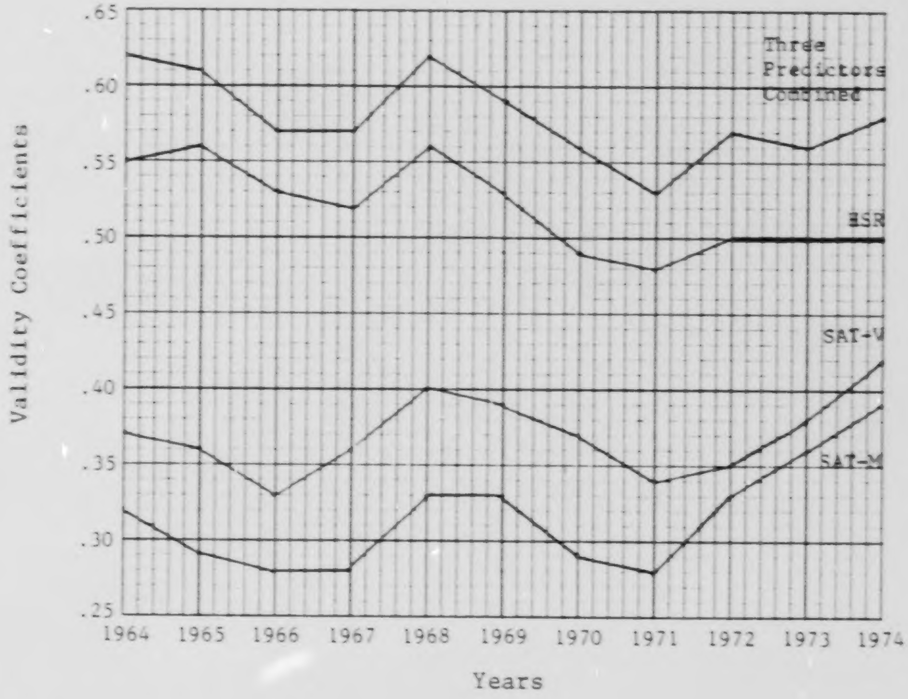


Figure 3. Combined Sex Sample Median Validity Trends 1964-1974



Comparison of SAT Score Trends in Selected Schools Judged to Have Traditional or Experimental Orientations

Rex Jackson
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

September 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

This study presents results of analyses of Scholastic Aptitude Test (SAT) scores for four graduating classes--1966, 1969, 1973, and 1976--in two contrasting groups of schools, one judged to be "traditional" and the other "experimental" in orientation. Since the national decline in mean SAT scores is attributed by some to increasing educational innovation in secondary schools, it was felt that an informal study of the kind reported here, while it could not support definitive conclusions, might provide some insight into whether further pursuit of this kind of inquiry would be fruitful.

All the schools in this study--16 in the traditional category and 12 in the experimental category--are public high schools. All but a few of the schools are highly regarded schools in affluent, suburban areas. Several are selective schools in urban areas. Mean SAT-verbal and SAT-mathematical scores for these schools for the years under study were consistently higher than the national averages--from 34 to 41 points for the traditional schools and from 47 to 60 points for the experimental schools. Within each group, the Eastern, Midwestern, and Western sections of the country are represented.

Schools were selected for the study and classified as "traditional" or "experimental" in general orientation with the assistance of staff members of the National Association of Secondary School Principals. In general, the schools nominated as experimental tended to be those that had adopted such innovations as minicourses and a range of elective courses, that had tried new methods in delivery of instruction, that had experimented with schedule modifications, and that in general were judged to provide more latitude for student choice. Thus, the experimental-traditional distinction does not reflect mainly differences in flexibility of course and program structure.

The resulting assignment of schools to the two groups was based on the judgment and knowledge of a single (but highly qualified) group of judges. It is likely that, had additional judgments been solicited, there would not have been complete agreement on all assignments to categories. However, there is no reason to doubt that, in the aggregate, the groups are contrasted on the intended dimension.

The original samples consisted of 19 traditional and 15 experimental schools. However, when the data were analyzed, it was found that several schools had either no students or very few students tested with the SAT for at least one of the years under study. These schools were dropped, reducing the samples to 16 traditional and 12 experimental schools.

The main result of this study is apparent from inspection of Table 1. From 1966 to 1976, mean scores for the traditional schools declined by 45 points on SAT-V and 24 points on SAT-M. For the experimental schools, the declines were 51 points on SAT-V and 30 points on SAT-M. For the national SAT candidate population (testing year data), the declines from 1965-66 through 1975-76 were 42 points on SAT-V and 26 points on SAT-M. Thus, the traditional schools show declines very close to the national figures, while the experimental schools show slighly larger declines.

Table 2 gives the percentage distributions of SAT scores for students in the two groups of schools for the years under study. It is apparent that the national experience of declining numbers of high scoring

Table 1. Mean SAT Scores for Students in the 1966, 1969, 1973, and 1976 Graduating Classes of Two Groups of Schools

	16 schools judged to have a traditional orientation			12 schools judged to have an experimental orientation		
Class	N	Mean SAT-V	Mean SAT-M	N	Mean SAT-V	Mean SAT-M
1966	5862	511	532	6092	531	549
1969	6342	525	532	7201	519	546
1973	6130	481	519	6491	493	529
1976	6079	466	508	6335	480	519

Table 2. Percentages of Students in Given SAT Score Ranges, Graduating Classes of 1966, 1969, 1973, and 1976 of Two Groups of Schools

(a) Verbal

	16 schools judged to have a traditional orientation				12 schools judged to have an experimental orientation			
Range	1966	1969	1973	1976	1966	1969	1973	1976
700-800	5.2	4.4	2.5	2.3	7.1	6.5	3.1	3.5
600-690	18.5	17.0	13.6	11.2	23.7	21.2	17.3	14.5
500-590	31.1	30.6	26.7	24.5	32.5	29.8	29.3	27.3
400-490	30.0	31.7	35.5	34.5	24.0	26.6	29.3	29.2
300-390	13.5	14.4	18.3	22.8	11.0	13.6	16.9	19.6
200-290	1.7	1.9	3.3	4.7	1.7	2.2	4.0	5.8

(b) Mathematical

	16 schools judged to have a traditional orientation				12 schools judged to have an experimental orientation			
Range	1966	1969	1973	1976	1966	1969	1973	1976
700-800	8.1	7.8	6.2	7.1	11.3	11.2	8.2	8.9
600-690	22.3	22.1	18.5	16.8	26.8	25.6	22.6	19.8
500-590	31.5	31.4	33.7	28.8	28.7	28.9	31.5	28.7
400-490	25.5	27.5	27.5	28.2	22.1	22.9	22.8	24.8
300-390	11.2	9.8	12.6	16.3	9.9	10.0	12.6	14.1
200-290	1.4	1.4	1.5	2.7	1.2	1.3	2.3	3.7

students is paralleled in these schools, as both groups show sharp drops between 1966 and 1976 in the percentage of students scoring over 600 on SAT-V or SAT-M. For the traditional schools the percentage scoring over 600 declined from 23.7 to 13.5 percent for SAT-V and from 30.4 to 23.9 percent for SAT-M. For the experimental schools the corresponding declines were from 30.8 to 18.0 percent for SAT-V and from 38.1 to 28.7 percent for SAT-M.

Tables A1 and A2 give the mean scores by year for the individual schools in the sample. Except for school XI, which exhibits the unusual pattern of a 16-point decline for mean SAT-V score and a 16-point increase for mean SAT-M score, all schools in the study show some decline. This result provides some reassurance with regard to the possible effect of classification errors. Any shifting of several schools from one category to the other as a result of a more elaborate classification procedure would not change the result much. The general finding is that the total group of schools in this study (predominantly public high schools in affluent, suburban areas) exhibits a pattern of SAT score decline very similar to the national pattern with very little discernible difference in the score experience of schools judged to have a more traditional orientation as contrasted with those judged to have an experimental orientation.

Since the schools judged to have an experimental orientation did show a slightly larger decline than the other schools or than students nationally, the question arises whether the results would tend to support at least to a small degree the notion that the score decline may be attributable in part to innovation. Because of the small size of the difference and because differences in score trends for the two groups of school could well have been produced by other factors (such as demographic changes in the areas served by the schools), this study appears to provide little support for the theory attributing score decline to innovation. By the same token, the study cannot be regarded as being conclusive in ruling this theory out. It should be noted that a fairly limited definition of "experimental" was used in defining that group, that the classification procedure was fallible, and (again) that the acknowledged influence of other factors on SAT score trends may have operated to obscure differences between these groups that would otherwise have appeared. At best, this exploratory study suggests that investigation of other theories about the SAT score decline have a greater likelihood of advancing understanding of the phenomenon.

Table A1. Numbers of Cases and Mean SAT-Verbal and SAT-Mathematical Scores, for Students from the 1966, 1969, 1973, and 1976 Graduating Classes of Selected Public High Schools Having a Traditional Orientation

	1966			1969			1973			1976			Differences in means (1966-1976)	
School	N	Mean SAT-V	Mean SAT-M	N	Mean SAT-V	Mean SAT-M	N	Mean SAT-V	Mean SAT-M	N	Mean SAT-V	Mean SAT-M	SAT-V	SAT-M
TA	346	492	527	466	499	530	420	478	529	462	466	523	26	4
TB	128	524	550	164	521	539	225	498	530	284	464	510	60	40
TC	259	554	570	277	530	542	221	503	548	247	520	558	34	12
TD	185	524	555	231	508	547	180	520	559	218	507	533	17	22
TE	501	509	530	557	496	519	545	482	508	410	473	508	36	22
TF	1055	510	522	820	516	536	781	491	507	751	480	507	30	15
TG	469	503	512	497	473	493	454	457	496	378	446	485	57	27
TH	58	422	461	57	449	472	55	424	464	76	418	449	4	12
TI	236	480	519	318	477	509	383	462	518	353	450	498	30	21
TJ	269	496	533	333	505	543	280	470	524	258	454	505	42	28
TK	228	497	519	325	501	518	324	474	508	351	454	503	43	16
TL	402	507	543	426	496	535	476	487	522	466	457	514	50	29
TM	449	503	524	489	495	536	500	472	525	466	444	497	59	27
TN	470	511	516	548	515	534	467	498	525	505	467	491	44	25
TO	256	505	520	245	506	530	475	463	514	456	461	504	44	16
TP	551	551	574	589	536	577	344	507	536	398	480	521	71	53

Table A2. Numbers of Cases and Mean SAT-Verbal and SAT-Mathematical Scores, for Students from the 1966, 1969, 1973, and 1976 Graduating Classes of Selected Public High Schools Having an Experimental Orientation

	1966			1969			1973			1976			Differences in means (1966-1976)	
School	N	Mean SAT-V	Mean SAT-M	N	Mean SAT-V	Mean SAT-M	N	Mean SAT-V	Mean SAT-M	N	Mean SAT-V	Mean SAT-M	SAT-V	SAT-M
XA	1082	498	516	1155	483	510	1084	461	499	988	453	487	45	29
XB	254	468	489	320	455	478	215	418	453	200	390	419	78	70
XC	758	533	554	928	524	553	737	500	552	758	478	525	55	29
XD	261	529	529	326	509	531	358	483	505	370	472	504	57	25
XE	333	529	539	388	501	510	346	491	526	426	468	504	61	35
XF	592	497	521	738	486	528	613	468	512	653	452	502	45	19
XG	640	510	532	773	510	534	759	474	492	739	471	500	39	32
XH	490	522	540	525	508	529	470	493	520	375	486	511	36	29
XI	152	520	547	342	533	569	344	503	557	281	504	563	16	(+16)
XJ	974	609	626	988	608	630	852	570	607	778	549	593	60	33
XK	254	576	611	288	546	629	253	534	605	288	543	581	33	30
XL	302	537	534	430	521	533	460	493	508	479	470	516	67	18

Correlations of SAT Scores with High School Record

Rex Jackson
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

January 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

The principal purpose of the Scholastic Aptitude Test (SAT) is to assist in predicting performance in college. Many studies have been performed of the effectiveness of the SAT (often in conjunction with some measure of high school performance) in fulfilling this purpose.

One result of examining the declines in mean scores of SAT candidates from 1963 to the present has been the suggestion that, although SAT scores continue to contribute to effective prediction of college grades, the SAT may have been overtaken by changes in high school programs that have made the test increasingly less related to the learning that takes place at the secondary school level. From this perspective, the principal message of score declines is not that student achievement has been adversely affected, but rather that more emphasis at the high school level has been placed on skills not measured by tests like the SAT.

One set of evidence bearing on this issue consists of studies of the statistical relationship of the SAT with measures of high school performance. If it is true that schools and teachers now give more emphasis to other kinds of skills and tend to reward more frequently achievement not dependent on the kinds of skills measured by the SAT, some decline in the level of this relationship would be expected.

The introduction of the Student Descriptive Questionnaire (SDQ) in 1971-72 permits analysis of the relationship of SAT scores and self-reported rank-in-class for those SAT candidates from each high school class responding to the SDQ. These correlations (based on 10 percent samples) are given in Table 1. There are several problems with the results in Table 1. First, the class ranks used in the analyses were self-reported ranks grouped into fairly coarse intervals. If actual ranks based on school transcripts had been available, the correlations would almost certainly have been higher. In addition, class ranks depend not only on individual performance but on that of classmates. Were a refined measure of class rank available, taking into account differences among schools, undoubtedly even higher correlations would result.

The absolute levels of the coefficients reported in Table 1 should be interpreted with the considerations noted above in mind. The trend over these five years, however, is unmistakably clear. For all practical purposes, the correlations of each part of the test with self-reported rank are unchanged across the years in the table. The slight variations from year-to-year are entirely inconsequential.

Data of the kind on which the analyses presented in Table 1 are based exist only for 1971-72 to the present. The decline in mean scores, however, began in 1963-64, and evidence on the relationship of SAT scores to high school performance over this longer period is also of interest. Certain relevant data can be found in studies of the effectiveness of SAT scores and high school record for predicting college grades. Since the mid-1960s many colleges have conducted such studies through the College Board Validity Study Service (VSS). Although attention is generally centered on the relationship of SAT scores and high school record (and combinations of these) with college grades, correlations of SAT scores with measures of high school performance are also regularly computed and reported as part of such studies. Table 2 gives the tenth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth percentiles of the

Table 1. Correlations of Self-Reported Rank in High School Class with SAT Scores for SAT Candidates from the 1971-72 through 1975-76 High School Classes Responding to the Student Descriptive Questionnaire

Class	SAT-V vs. rank	SAT-M vs. rank
1971-72	.493	.518
1972-73	.501	.520
1973-74	.505	.536
1974-75	.488	.529
1975-76	.498	.536

Notes:
1. Rank was coded as follows: 5 = top tenth, 4 = second tenth, 3 = second fifth, 2 = third fifth, 1 = bottom two fifths.
2. Analyses for each class were based on ten percent samples of SAT candidates responding to the SDQ.
3. For the 1971-72 class, the sample was limited to students participating as seniors.

distribution of correlation coefficients of SAT scores with measures of high school record for a number of validity studies based on students who entered VSS-participating colleges in years from 1964 through 1974.

A college conducting a validity study through VSS generally requests analyses based on a number of student subgroups. The summary in Table 2 is limited to subgroups classified as "all students" or "all SAT-takers" or "liberal arts students" (if the previous categories were not used). If a given study provided results for males and females combined (for one of the above noted groups), the correlations based on this group were entered into the distributions. If only separate sex analyses were available, the separate results for males and for females were included.

The correlations summarized in Table 2 are, in general, based on analyses using better measures of high school record than those for the analyses reported in Table 1. Generally a less coarse scale is used and, in all cases, the numbers are based on high school transcripts rather than student-reported ranks. However, the samples of students on which the correlations summarized in Table 2 are computed, are selected groups. They are (in virtually all cases) groups selected for college acceptance jointly on SAT scores and high school record. It is well known that, for such selected groups, intercorrelations of variables used in selection (and outcome variables as well) will be lower than the corresponding correlations for groups not subject to such selection. The differences between the levels of correlation for unselected groups in Table 1 and the median correlations for the select groups in Table 2 are largely attributable to these selection effects.

Table 2. Distributions of Correlations Between High School Record (Rank or GPA) and SAT Scores for Selected Groups of Students in Colleges Participating in the College Board Validity Study Service (1964-1974)

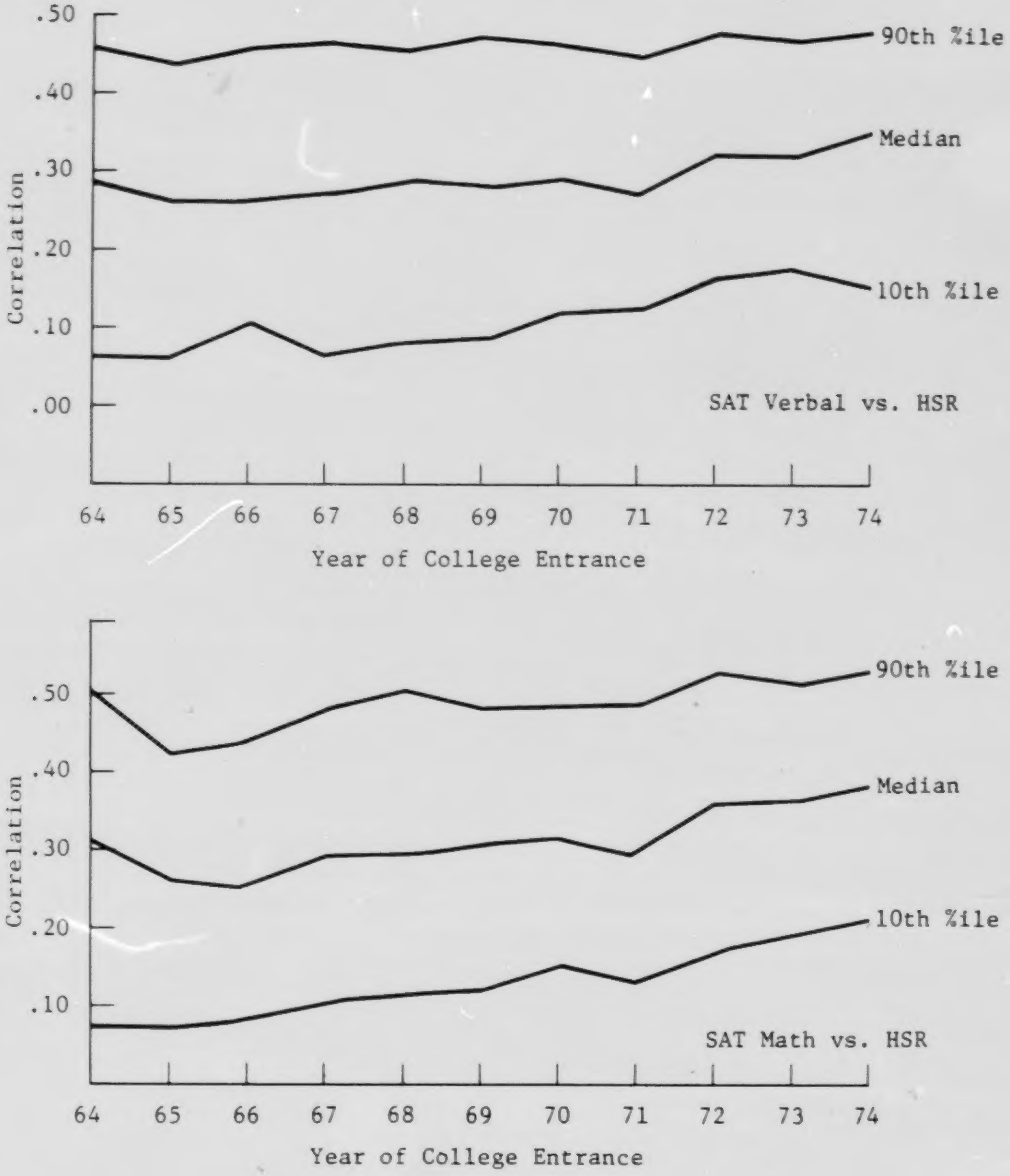
Correlations with SAT-Verbal

Percentile	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	Percentile
90	.460	.441	.460	.463	.456	.475	.465	.450	.482	.477	.484	90
75	.386	.371	.379	.392	.396	.383	.380	.360	.423	.410	.431	75
50	.288	.262	.264	.274	.290	.287	.293	.276	.328	.322	.354	50
25	.184	.148	.166	.165	.174	.201	.209	.208	.225	.235	.256	25
10	.066	.064	.108	.062	.081	.085	.122	.124	.165	.178	.159	10

Correlations with SAT-Mathematical

Percentile	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	Percentile
90	.504	.427	.440	.480	.506	.465	.483	.486	.530	.514	.532	90
75	.402	.358	.363	.393	.400	.383	.414	.386	.463	.450	.470	75
50	.318	.263	.253	.291	.297	.308	.314	.299	.360	.363	.389	50
25	.203	.167	.175	.168	.191	.217	.229	.213	.258	.280	.308	25
10	.075	.076	.086	.105	.114	.127	.151	.134	.171	.198	.214	10
Number of groups	131	146	185	200	119	142	271	99	172	136	154	

Figure 1. Selected Percentiles of the Distribution of Correlations of SAT Scores with High School Record for CEEB Validity Study Groups



Inspection of the trends in median correlations (and other percentiles) in Table 2 lends no support to the notion that either SAT verbal or SAT mathematical scores have become less related to high school grades over the period covered by the table. These trends are also graphed in Figure 1. If anything, SAT scores and high school grades appear to be more closely related in more recent years. Because the results reported in Table 2 are substantially influenced by college selectivity, it is possible that the year-to-year differences (and, in particular, the apparently higher median correlations in more recent years) are attributable to changes in the characteristics of colleges participating in VSS over the years studied. In the absence of more detailed information on this point, it would be premature to infer that SAT scores have become increasingly more related to high school performance as measured by rank-in-class or grade point average. Clearly, however, the available data do not directly support the conclusion that high school evaluations of student achievement have become progressively less related to scores on the SAT.

An Examination of Declining Numbers of High-Scoring SAT Candidates

Rex Jackson
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

October 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

One of the more mysterious aspects of the decline in verbal and mathematical scores by Scholastic Aptitude Test (SAT) candidates has been the decline in numbers of high scoring students. Had the number of high scoring students not declined, a theory attributing declines in mean scores to an enlargement of opportunity for college attendance and consequent changes in the composition of the SAT candidate group would seem more plausible. Since the declines in mean SAT scores are not attributable solely to increased numbers of low scoring students, but also to smaller numbers of high scoring students, it appears that such a theory cannot fully explain the decline in scores. The purpose of this paper is to provide some detailed information on the decline in numbers of high scoring students in the SAT candidate group and to examine several explanations of this phenomenon that have been proposed.

Table 1 gives the distributions of scores for students grouped by testing year. If a "high score" is taken to be one of 600 or greater, it is apparent that the percentage of test takers attaining high scores declined fairly sharply for the verbal test and more gradually for the mathematical test during the ten years shown in the table. However, it is not true that numbers of high scoring students declined steadily throughout this period.

For the verbal test, the number of students scoring over 600 actually increased from 1966-67 to 1967-68 and then remained almost level through 1969-70. From 1969-70 to 1970-71, this number dropped from 189,300 to 170,800--a drop of nearly 20,000--and continued to fall by about another 20,000 to 150,700 in 1971-72; by about 22,000 to 128,200 in 1972-73; by another 7,000 to 121,100 in 1973-74; by 11,000 to 110, 300, in 1974-75; and by about 2,000 more to 108,200 in 1975-76. From a high of 189,300 verbal scores of 600 or higher attained by students taking the test in 1969-70, the number of such scores fell by more than 80,000 to 108,200 in 1975-76.

For the mathematical test the pattern is similar - an increase in high scores from 1966-67 to 1967-68; almost constant numbers from 1967-68 through 1970-71; then a decline from 278,600 scores over 600 in 1970-71 to 213,900 in 1975-76--a decline of almost 65,000. The similarity of this pattern to that for the verbal test is not exact - the sharp decline in high verbal scores began in 1970-71, while that for mathematical scores began in 1971-72. It is clear, however, that declining numbers of high scores have not been characteristic of the entire period of score decline--from 1962-63 to the present--but only of the last five or six years.

Table 2 gives means, standard deviations, and numbers of cases for students grouped on a testing year basis for the last 20 years. Several facts emerge from examining this table and the data presented earlier:

(a) The total mean score declines from 1962-63 to 1969-70 were only 18 points (478-460) for SAT-V and 14 points (502-488) for SAT-M, average annual declines of 2.6 and 2.0 points, respectively. These contrast with declines from 1969-70 to 1975-76 of 31 points (460-429) for SAT-V and 18 points (488-470) for SAT-M, average annual declines of 5.2 and 3.0 points, respectively, over this six year period.

Table 1. Numbers (in thousands) and Percentages of Candidates in Given Score Ranges on the Scholastic Aptitude Test, 1966-67 through 1975-76

(a) Verbal

Score range	1966-67		1967-68		1968-69		1969-70		1970-71	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
700-800. . . .	33.2	(2.3)	30.5	(2.0)	27.9	(1.8)	27.6	(1.7)	25.6	(1.7)
600-699. . . .	145.3	(10.2)	156.7	(10.1)	157.0	(9.9)	161.7	(10.1)	145.2	(9.4)
500-599. . . .	339.4	(23.9)	360.8	(23.4)	379.3	(23.9)	378.4	(23.6)	361.8	(23.5)
400-499. . . .	503.5	(35.4)	534.9	(34.6)	548.8	(34.6)	530.9	(33.1)	520.2	(33.8)
300-399. . . .	325.4	(22.9)	378.1	(24.5)	377.7	(23.8)	401.7	(25.0)	385.7	(25.1)
200-299. . . .	75.7	(5.3)	82.9	(5.4)	94.9	(6.0)	105.4	(6.6)	98.6	(6.4)
	1422.5		1543.8		1585.6		1605.9		1537.2	

Score range	1971-72		1972-73		1973-74		1974-75		1975-76	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
700-800. . . .	20.0	(1.4)	16.0	(1.1)	16.8	(1.2)	14.8	(1.1)	15.9	(1.1)
600-699. . . .	130.7	(9.0)	112.2	(8.0)	104.3	(7.7)	95.5	(7.0)	92.3	(6.5)
500-599. . . .	332.5	(22.8)	300.1	(21.5)	279.6	(20.6)	289.4	(21.1)	273.1	(19.3)
400-499. . . .	493.4	(33.8)	470.2	(33.6)	463.7	(34.2)	468.5	(34.2)	474.2	(33.5)
300-399. . . .	378.6	(25.9)	390.7	(27.9)	378.5	(28.0)	376.4	(27.4)	405.8	(28.7)
200-299. . . .	104.7	(7.2)	109.2	(7.8)	111.2	(8.2)	126.7	(9.2)	153.6	(10.9)
	1459.9		1398.4		1354.0		1371.2		1415.0	

(Continued)

Table 1 (Continued)

(b) Mathematical

Score range	1966-67 Number	1966-67 Percent	1967-68 Number	1967-68 Percent	1968-69 Number	1968-69 Percent	1969-70 Number	1969-70 Percent	1970-71 Number	1970-71 Percent
700-800.	55.5	(3.9)	56.2	(3.6)	55.9	(3.5)	56.0	(3.5)	55.7	(3.6)
600-699.	199.0	(14.0)	225.9	(14.6)	224.0	(14.1)	224.4	(14.0)	222.9	(14.5)
500-599.	413.3	(29.1)	442.2	(28.6)	454.1	(28.6)	457.7	(28.5)	428.3	(27.9)
400-499.	465.8	(32.7)	478.5	(31.0)	495.1	(31.2)	497.6	(31.0)	483.2	(31.4)
300-399.	248.1	(17.4)	282.4	(18.3)	296.3	(18.7)	303.5	(18.9)	291.9	(19.0)
200-299.	40.7	(2.9)	58.6	(3.8)	60.2	(3.8)	66.6	(4.1)	55.2	(3.6)
	1422.5		1543.8		1585.6		1605.9		1537.2	

Score range	1971-72 Number	1971-72 Percent	1972-73 Number	1972-73 Percent	1974-74 Number	1974-74 Percent	1974-75 Number	1974-75 Percent	1975-76 Number	1975-76 Percent
700-800.	44.3	(3.0)	44.8	(3.2)	46.3	(3.4)	45.5	(3.3)	44.1	(3.1)
600-699.	203.2	(13.9)	184.0	(13.2)	175.9	(13.0)	168.7	(12.3)	169.8	(12.0)
500-599.	415.3	(28.4)	402.8	(28.8)	365.2	(27.0)	360.6	(26.3)	371.2	(26.2)
400-499.	449.4	(30.8)	429.1	(30.7)	413.1	(30.5)	413.1	(30.1)	425.6	(30.1)
300-399.	290.2	(19.9)	280.8	(20.1)	291.6	(21.5)	315.1	(23.0)	314.3	(22.2)
200-299.	57.6	(3.9)	57.0	(4.1)	61.9	(4.6)	68.1	(5.0)	89.9	(6.4)
	1459.9		1398.4		1354.0		1371.2		1415.0	

(b) For the years from 1962-63 to 1969-70 the relatively small annual declines in mean scores were accompanied by increasing numbers of students taking the test--from 933,100 in 1962-63 to 1,605,900 in 1969-70.

(c) As indicated earlier, although the percentages of high scoring students declined as the size of the test-taking groups increased, the absolute numbers of such students did not decline prior to 1969-70. Thus, for the first seven years, the declines in mean scores can be attributed to increasing numbers of students in the middle and lower score ranges, and not to smaller numbers of students scoring over 600.

While the data presented here cannot support any definite conclusions about the validity of a theory attributing score declines to changes in the SAT test taking group, the SAT data for the years from 1962-63 through 1969-70 seem consistent with such a theory. A considerable body of evidence that could be brought to bear on this issue will be available to the panel in other reports. However, a few salient points may be noted here:

Table 2. Means, Standard Deviations, and Numbers of Cases* for all Candidates Taking the SAT, by Year, 1956-57 through 1975-76

Testing	SAT-Verbal		SAT-Mathematical		
year	Mean	S.D.	Mean	S.D.	N/1000
1956-57.	473	105	496	111	270.5
1957-58.	472	106	496	109	376.8
1958-59.	475	106	498	110	469.7
1959-60.	477	108	498	109	564.2
1960-61.	474	108	495	109	716.3
1961-62.	473	109	498	112	802.5
1962-63.	478	109	502	111	933.1
1963-64.	475	108	498	113	1,163.9
1964-65.	473	108	496	112	1,361.2
1965-66.	471	108	496	111	1,381.4
1966-67.	467	109	495	110	1,422.5
1967-68.	466	108	494	112	1,543.8
1968-69.	462	108	491	111	1,585.6
1969-70.	460	110	488	113	1,605.9
1970-71.	454	109	487	113	1,537.2
1971-72.	450	108	482	112	1,459.9
1972-73.	443	107	481	112	1,398.4
1973-74.	440	107	478	115	1,354.0
1974-75.	437	108	473	114	1,371.2
1975-76.	429	109	470	118	1,415.0

*The term "case" denotes attendance at an SAT administration. Thus, if an individual was tested twice, he (she) is counted twice in these tabulations.

(a) The Condition of Education (1976), published by NCES, gives the following figures for ratios of first-time degree credit enrollment to numbers of high school graduates one year earlier (Table 4.13):

School year	Graduates (in thousands)	School year	College entrance (in thousands)	Percent of graduates
1957-58	1,506	1958-59	772	51.3
1962-63	1,957	1963-64	1,046	53.6
1964-65	2,665	1965-66	1,442	54.1
1967-68	2,702	1968-69	1,630	60.3
1970-71	2,943	1971-72	1,766	60.0
1973-74	3,069	1974-75	1,854	60.4

It appears that there was a sharp rise in percentage of high school graduates attending college during the 1960's and a leveling off since then at about 60 percent.

(b) The number of students participating in the ACT assessment program shows a trend similar to that for the SAT--rising from 1962-63 to 1970-71 and then declining for the next few succeeding years. (Source: Leo Munday, "Declining Admissions Test Scores," ACT Research Report No. 71.)

(c) An analysis of distribution of SAT score reports by W. B. Schrader ("Distribution of SAT Scores to Colleges as an Indicator of Changes in the SAT Candidate Population," September 1976) indicates that the overall number of SAT score reports sent to colleges almost doubled between 1960-61 and 1966-67 (consistent with the increase in number of candidates). In addition, there were marked increases in the percentages of reports to comprehensive universities and colleges and to two-year colleges and decreases in percentages (but not number) of reports to the most selective institutions--private research universities and liberal arts colleges. Between 1966-67 and 1973-74 both the total number of reports and the percentage distribution of reports over types of colleges were much more similar.

The three sets of evidence alluded to above and the SAT data discussed earlier that changes in the numbers of students going to college (and probably in the composition of that population) during the early years of the score decline may have caused progressive changes in the SAT candidate groups that could help to account for this part of the decline. For the years from about 1970-71 to the present, there appears to be less evidence consistent with a population shift hypothesis and there is a need, particularly for the last six years, to explain a decrease in the numbers of high scoring students taking the SAT (not just an increase in numbers of lower scoring students).

The remainder of this paper will address several explanations that have been offered for the decline in high scoring students. Before examining these issues, it may be useful to provide assurance that the declines in numbers of high scoring students are not simply the result of less repeated test taking. Since the distributions presented earlier are based on all students tested in a given year, it is conceivable that the smaller numbers of over-600 scores are caused by a tendency of high-scoring students in more recent years to take the test only once. Since 1971-72, distributions of scores have been calculated also for so-called "cohorts"--groups of students aggregated according to their year of high school graduation. For these distributions only one set of scores (the most recent one) is used for each student. The cohort distributions of SAT scores are given in Table 3.

It is apparent from Table 3 that it is not just the numbers of SAT scores over 600 that have declined. The number of students taking the SAT and attaining these scores has also declined. The following table of numbers (in thousands) of over-600 scores for students grouped in the two ways is informative:

	Verbal		Mathematical	
	Test year	Cohort	Test year	Cohort
(a) 1971-72.	150.7	116.6	247.5	182.6
(b) 1975-76.	108.2	82.0	213.9	163.4
(b)/(a)	.72	.70	.86	.89

It is clear that both sets of data show similar rates of decline in numbers of high scores over the last five years.

The explanations of the declining numbers of high scoring students generally fall in two classes: (a) those that assume that there are simply fewer highly able students in the high school graduate populations of recent years, and (2) those that assume there are as many highly able high school students, but that fewer are taking the SAT. Data relating directly to hypotheses in the first class are hard to come by. However, examination of certain existing data may help in deciding how much credence should be given to explanations of the second type. Among these are: (1) the assertion that many of the able high school students are not attending college and thus not taking any admissions test, (2) the assertion that able students are being admitted to college on the basis of other information (including PSAT/NMSQT scores) and thus take the SAT less frequently, and (3) an assertion that able students are less likely to take both the SAT and the ACT Assessment and thus appear less often in the SAT statistics. (This last theory could, in fact, explain a loss of high scoring students in both programs; ACT reports a constant percentage of high scoring students--14 percent in the 26-36 interval--over the last five years which in the light of a lowering of total number of students tested over this period would suggest some decline in numbers of high scoring students.)

One set of evidence that may relate to all the assertions described above is the body of information developed from the responses to the Student Descriptive Questionnaire (SDQ). The first cohort for which complete SDQ information is available is 1972-73 and the most recent is 1975-1976. From Table 3 it can be seen that from the 1972-73 to 1975-76 cohort, the number of students scoring over 600 declined from 98,256 (9.7%) to 81,964 (8.2%) for the verbal test and from 169,029 (16.7%) to 163,398 (16.3%) for the mathematical test. Over this same period, examination of SDQ responses for 10% samples of SAT takers reveals that: the percent of students reporting they ranked in the top 10th of their

Table 3. Numbers (and Percentages) of Candidates in Given Score Ranges on the Scholastic Aptitude Test, 1971-72 through 1975-76, for Students Grouped According to Year of High School Graduation

(a) Verbal

Score range	1971-72		1972-73		1973-74		1974-75		1975-76	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
700-800. .	17,560	(1.7)	12,537	(1.2)	14,677	(1.5)	11,380	(1.1)	12,288	(1.2)
600-690. .	99,070	(9.7)	85,719	(8.4)	84,089	(8.5)	67,753	(6.8)	69,676	(7.0)
500-590. .	239,755	(23.4)	221,295	(21.8)	212,480	(21.6)	202,833	(20.4)	199,099	(19.9)
400-490. .	334,118	(32.7)	341,880	(33.7)	325,516	(33.0)	338,015	(33.9)	321,520	(32.2)
300-390. .	261,233	(25.5)	269,510	(26.6)	267,256	(27.1)	278,311	(27.9)	289,798	(29.0)
200-290. .	71,084	(6.9)	83,912	(8.3)	81,221	(8.2)	98,136	(9.8)	107,428	(10.7)
	1,022,820		1,014,853		985,239		996,428		999,809	

(b) Mathematical

Score range	1971-72		1972-73		1973-74		1974-75		1975-76	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
700-800. .	37,067	(3.6)	31,588	(3.1)	36,242	(3.7)	33,005	(3.3)	37,511	(3.8)
600-690. .	145,535	(14.2)	137,441	(13.5)	133,602	(13.6)	122,511	(12.3)	125,887	(12.6)
500-590. .	284,242	(27.8)	291,991	(28.8)	262,211	(26.6)	257,283	(25.8)	254,672	(25.5)
400-490. .	311,713	(30.5)	306,916	(30.2)	302,225	(30.7)	299,528	(30.1)	290,467	(29.1)
300-390. .	200,502	(19.6)	202,317	(19.9)	209,452	(21.3)	230,912	(23.2)	227,729	(22.8)
200-290. .	43,621	(4.3)	44,451	(4.4)	41,383	(4.2)	53,152	(5.3)	63,510	(6.4)
	1,022,680		1,014,704		985,115		996,391		999,776	

class rose from 15.3% to 17.8%; the percent with average reported grades of 3.75--4.00 rose from 11.5% to 13.1%; the percent reporting having received 3 or more scholastic honors rose from 13.8% to 16.3%; the percent planning graduate study rose from 26.8% to 31.0%. If it is true that there are as many able students among high school graduates but that fewer are currently taking the SAT (for whatever reason), one would expect the withdrawal of these students from the SAT pool to be reflected in declines rather than increases in the percentage of students giving these responses that are more characteristic of high scoring students.

Despite the evidence from the SDQ that the SAT taking group has not changed in recent years in ways that would help to account for smaller numbers and percentages of high scoring students, some analysis of the possible effects of changes in test-taking patterns on the composition of the SAT group may be warranted. In the case of the PSAT to SAT comparison it would be ideal to have analyses of matched and unmatched candidates over time, so that any trend among high scoring students to take only the PSAT could be ascertained. Unfortunately these analyses do not exist and cannot be done now since the necessary PSAT data is not available for the early years. However, a match of records for high school juniors taking the PSAT in October, 1973 with SAT files yields the following partition of that group.

	Verbal			Mathematical		
	PSAT only	PSAT and SAT	Total	PSAT only	PSAT and SAT	Total
PSAT 60-80	12,183	49,800	61,983	31,889	99,661	131,550
Score 20-59	424,773	580,845	1,005,618	404,789	530,984	935,773
	436,956	630,645	1,067,601	436,678	630,645	1,067,323

It is apparent that students attaining high scores on the PSAT are much more likely than lower scoring students to take the SAT also. For the verbal test 80 percent of the high scoring PSAT candidates also took the SAT while only 58 percent of the lower scoring candidates did. For the mathematical test the corresponding percentages were 76 percent and 57 percent.

Another way of considering this matter is to observe from Table 3 that for the verbal test the number of students scoring 600 or higher on the SAT fell from 116,630 for the 1971-72 cohort to 79,133 for the 1974-75 cohort. (The 1974-75 cohort is the one that includes the 1973 PSAT juniors.) Thus, almost 37,500 fewer high scoring SAT candidates have to be accounted for. Since there were only about 12,000 students from the 1974-75 class who took the PSAT as juniors gaining verbal scores of 60-80 but not repeating the SAT, it does not appear that many of the "missing" SAT candidates can be found taking the PSAT only.

This analysis is (of necessity) a fairly rough one. Some of the students scoring lower than 60 on the PSAT and who did not take the SAT would have obtained scores higher than 600 on the SAT. Similarly, some of those scoring 60 or higher on the PSAT would have scored lower than 600 on the SAT. However, available data on change in scores from junior year PSAT scores to later SAT scores suggests that even if all the 1973 PSAT juniors had later taken the SAT (an extreme assumption), the number of students in the 1974-75 cohort scoring 600 or higher on the SAT would not have increased by much more than 12,000. It may be that such factors as early college entrance or waiving by colleges of SAT requirements has had some effect on numbers of high scoring SAT candidates, but available data do not indicate that acceptance of PSAT scores in place of SAT scores (if this is occurring) has had a large effect.

The other suggested explanation relating to test taking pattern attributes declines in numbers of high scoring students to less joint taking of SAT and the ACT Assessment. Again, ideal data for testing this theory (counts of matched and unmatched SAT and ACT candidates by score level over time) do not currently exist, although the necessary raw data is undoubtedly available. In the absence of such counts, however, the following analysis may shed some light on the question.

In certain states nearly all college-bound students take the SAT and very few take ACT; in certain other states the situation is reversed; and in some states both tests are taken by sizable numbers of students. If the theory in question is correct, we should see relatively little decline in numbers of high scoring students in those states where very few students take ACT, since the effect of a shift to ACT or of less joint taking of SAT and ACT could not be sizable in such states. Because the data were readily available in convenient form, the numbers of students scoring 600 or higher on SAT-V and SAT-M were tabulated for the 1972-73 and 1975-76 cohorts for 18 states that were known to be states in which students taking SAT far outnumber those taking ACT. the results are:

SAT	Verbal		Mathematical	
score	1972-73	1975-76	1972-73	1975-76
600-800.	52,875 (8.6%)	45,135 (7.1%)	91,808 (14.9%)	90,588 (14.3%)
200-590.	563,619 (91.4%)	588,080 (92.9%)	524,582 (85.1%)	542,598 (85.7%)
	616,494	633,215	616,390	633,186

It is clear that both numbers and percentage of students scoring over 600 are declining in these states as well as in the rest of the nation. The hypothesized ACT effect probably cannot fully explain this decline. However, it may be noted that the rate of decline was slightly less steep in these states than for the entire nation. For the verbal test, for example, there were 7,740 fewer students in the 1975-76 cohort from these states scoring over 600. For the total groups, the decline was 16,292. Thus despite the fact that more than 60 percent of the SAT takers from each cohort are from these states, more than half the decline in numbers of students scoring over 600 occurred elsewhere.

For a variety of reasons, these data cannot support definitive conclusions. However, the results suggest that the decline in numbers of high scoring students is not solely attributable to changing test-taking patterns. The data are consistent with the hypothesis that smaller numbers of high scoring students may be due in part to a lower rate of participation in both testing programs by such students, but they cannot confirm that hypothesis. Other studies would be needed to provide such confirmation.

Summary

1. Although percentages of students scoring over 600 on SAT-V and SAT-M have been declining at least since 1966-67, the absolute numbers of such students began to decline only in 1970-71.

2. Data on trends for numbers of students scoring in given ranges on the SAT as well as overall numbers of SAT participants appear to be consistent with the theory that declines prior to 1970-71 can be accounted for largely by changes in the SAT taking group.

3. Declines in absolute numbers of high scoring students for the last six years accompanied by increases in percentage of SAT takers reporting high levels of high school achievement make the population shift theory less plausible as an explanation of score declines over this period.

4. Limited data suggest that changes in test taking patterns (for example, reduced joint taking of PSAT and SAT or ACT and SAT) could account for some of the decline in numbers of high scoring students, but further study would be needed before confident statements on this topic could be made. Data on the composition of the SAT group collected by means of the SDQ (as noted in point three above), data on matched PSAT and SAT student records, and data on SAT takers in selected states all suggest that the declining numbers of high scoring students in recent years are not solely attributable to changes in the composition of the test taking group (brought about by changing patterns of test taking or by other factors).

Mean Scores for PSAT and PSAT/NMSQT Junior Candidates

Rex Jackson
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

December 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

A table showing numbers of cases, means, and standard deviations for PSAT (or PSAT/NMSQT) junior candidates for the years 1959-60 through 1976-77 is attached. Also attached are graphs of trends in PSAT mean scores and SAT testing year mean scores. (Since we have results from the October 1976 administration of the PSAT/NMSQT -- given once a year -- we can show the 1976-77 means for that test. The 1976-77 SAT testing year, however, is not complete.)

Over the years, the PSAT has been taken primarily by juniors although some seniors and sophomores also take the test. Currently about 95 percent of all candidates are juniors. Because the mix of seniors and sophomores in the total candidate group and their characteristics have changed for a variety of reasons (including use of the PSAT for different scholarship selection purposes), the data presented here are based on junior candidates only -- the group of PSAT candidates presumed to have been most stable in its composition over time.

Data for the first few years of the PSAT program (when the program was new and numbers of students taking the test were increasing rapidly) should probably be disregarded. Since 1963-64 (a high point for PSAT verbal mean scores and one year after the SAT high point), PSAT verbal means have declined in most years. From a high of 43.3 (equivalent to 433 on the SAT scale), PSAT verbal means have declined to a low of 40.5 in 1976-77.

Thus, although the PSAT pattern is not as regular as that for the SAT, the PSAT verbal means also exhibit some decline. It should also be noted that the PSAT and the National Merit Scholarship Qualifying Test were merged in 1971-72 (to become the PSAT/NMSQT). Thus, the increases in mean scores in 1971-72 and 1972-73 may be caused, at least in part, by a change in the testing program that resulted in an increase in high scoring candidates. Without this change the trends in mean scores for SAT-V and PSAT-V might have been even more nearly parallel.

The trend for PSAT mathematical mean scores does not appear to parallel the gradual decline exhibited by the SAT mathematical means. However, the possibility that the merger of the PSAT and the NMSQT has had an effect on mean scores for more recent years is also evident in the graph of PSAT mathematical means.

Some evidence concerning the possible effects of the merger may be found in the data on numbers of participating schools and students for the years just preceding and just following the merger.

	Year	Number of schools	All candidates	Average number of candidates per school
PSAT.	1968-69	15,522	1,284,285	82.7
	1969-70	14,883	1,162,654	78.1
	1970-71	14,398	1,133,839	78.8
PSAT/NMSQT.	1971-72	16,936	1,161,296	68.6
	1972-73	17,354	1,117,556	64.4
	1973-74	17,485	1,134,635	64.9

(Note that there are figures for all candidates, not just junior candidates.)

Summary Statistics of PSAT Scores for Juniors

	Male			Female			Total		
Junior year	N	M	SD	N	M	SD	N	M	SD
PSAT-Verbal									
1959-60 . . .	186,791	41.0	11.2	191,365	41.3	10.8	378,156	41.2	11.0
1960-61 . . .	229,781	40.9	10.7	232,008	41.0	10.4	461,789	40.9	10.6
1961-62 . . .	253,001	42.1	10.7	254,839	42.6	10.6	507,839	42.3	10.7
1962-63 . . .	317,085	42.8	10.7	330,624	42.9	10.5	647,709	42.9	10.5
1963-64 . . .	379,282	43.1	10.8	395,793	43.4	10.6	775,075	43.3	10.7
1964-65 . . .	402,465	42.7	10.6	421,496	43.0	10.4	823,961	42.9	10.5
1965-66 . . .	447,808	42.5	11.0	477,685	42.0	10.9	926,035	42.2	10.9
1966-67 . . .	461,994	42.7	11.0	496,478	42.7	10.7	958,974	42.7	10.8
1967-68 . . .	493,831	42.0	10.8	538,421	42.1	10.7	1,032,945	42.0	10.8
1968-69 . . .	508,888	42.2	11.5	555,304	42.9	11.6	1,064,986	42.6	11.6
1969-70 . . .	507,570	42.2	10.8	559,721	42.2	10.8	1,067,834	42.2	10.8
1970-71 . . .	492,990	41.4	11.0	553,583	41.4	10.7	1,047,135	41.4	10.8
*1971-72 . . .	512,622	42.1	10.6	576,393	42.4	10.5	1,090,868	42.2	10.6
*1972-73 . . .	491,385	42.6	11.1	564,975	42.7	10.9	1,056,752	42.7	11.0
*1973-74 . . .	492,683	42.0	10.8	574,424	41.7	10.7	1,067,601	41.8	10.8
*1974-75 . . .	503,959	42.2	10.7	589,840	41.1	10.4	1,094,328	41.6	10.5
*1975-76 . . .	508,834	41.8	11.1	602,818	40.4	11.1	1,112,483	41.0	11.1
*1976-77 . . .	510,543	40.9	10.4	609,174	40.1	10.3	1,120,542	40.5	10.4

Summary Statistics of PSAT Scores for Juniors (Continued)

Junior Year	Male N	Male M	Male SD	Female N	Female M	Female SD	Total N	Total M	Total SD
PSAT-Mathematical									
1959-60	186,791	47.6	10.2	191,365	42.4	8.6	378,156	45.0	9.8
1960-61	229,755	47.5	10.2	231,956	42.3	9.2	461,711	44.8	10.1
1961-62	252,995	48.6	10.4	254,820	43.6	9.4	507,804	46.1	10.2
1962-63	317,069	48.7	10.6	330,599	44.3	9.7	647,668	46.5	10.4
1963-64	379,282	47.6	10.5	395,793	43.5	9.3	775,075	45.6	10.1
1964-65	402,465	46.9	10.9	421,496	42.6	9.8	823,961	44.7	10.6
1965-66	447,803	48.3	11.1	477,678	43.8	10.2	926,023	45.9	10.9
1966-67	461,986	47.1	11.5	496,455	43.1	10.7	958,943	45.0	11.3
1967-68	493,829	46.4	10.8	538,418	42.7	9.7	1,032,940	44.5	10.4
1968-69	508,858	47.4	11.5	555,260	44.0	10.5	1,064,912	45.6	11.1
1969-70	507,557	47.0	11.4	559,714	43.2	10.6	1,067,814	45.0	11.1
1970-71	492,982	48.4	11.5	553,568	44.1	10.6	1,047,112	46.1	11.2
*1971-72	512,622	47.3	11.6	576,393	43.4	10.5	1,090,868	45.2	11.2
*1972-73	491,294	49.1	11.4	564,864	45.0	10.4	1,056,550	46.9	11.1
*1973-74	492,573	47.8	11.8	574,256	43.5	10.9	1,067,323	45.5	11.5
*1974-75	503,790	48.2	11.2	589,620	43.9	9.9	1,093,938	45.9	10.7
*1975-76	508,700	47.9	11.4	602,631	43.5	10.4	1,112,159	45.5	11.1
*1976-77	510,435	47.3	11.5	609,036	43.0	10.4	1,120,298	45.0	11.1

*PSAT/NMSQT

The increase in school participation is striking. After slight declines from 1968-69 to 1969-70 and 1969-70 to 1970-71, the number of schools increased in 1971-72 by 2,538 and in 1972-73 by another 418 -- a total of 2,956. At the same time both total candidate volume and number of junior candidates (see attached table) increased only slightly between 1970-71 and 1971-72 and actually declined from 1971-72 to 1972-73.

We know that volumes in schools that traditionally participated in the PSAT program were declining slightly (see the 1968-69 to 1970-71 figures), and the failure of volumes to increase significantly in 1971-72 and 1972-73 can probably be attributed partly to a continuation of this trend. However, the fact that the increase of nearly 3,000 schools (more than a 20 percent increase) from 1970-71 to 1972-73 was not accompanied by a larger number of students tested almost surely means that many of these new schools tested only a few students each. Undoubtedly, these students tended to be the ones judged by the schools to be best able to compete in the National Merit program. Given the signs of declines in numbers of students tested in traditional PSAT schools (possibly relatively low scoring students) and the replacement or partial replacement of these students by high scoring students in schools attracted to the test by the National Merit connection, there are substantial reasons for believing that sharp increases in scores in the years immediately after the merger were due to changes in the composition of the candidate group. At least for the verbal test, a discounting of the 1971-72 and 1972-73 increases in mean PSAT/NMSQT scores would result in very nearly parallel trend lines for the SAT-V and PSAT-V.

Figure 1

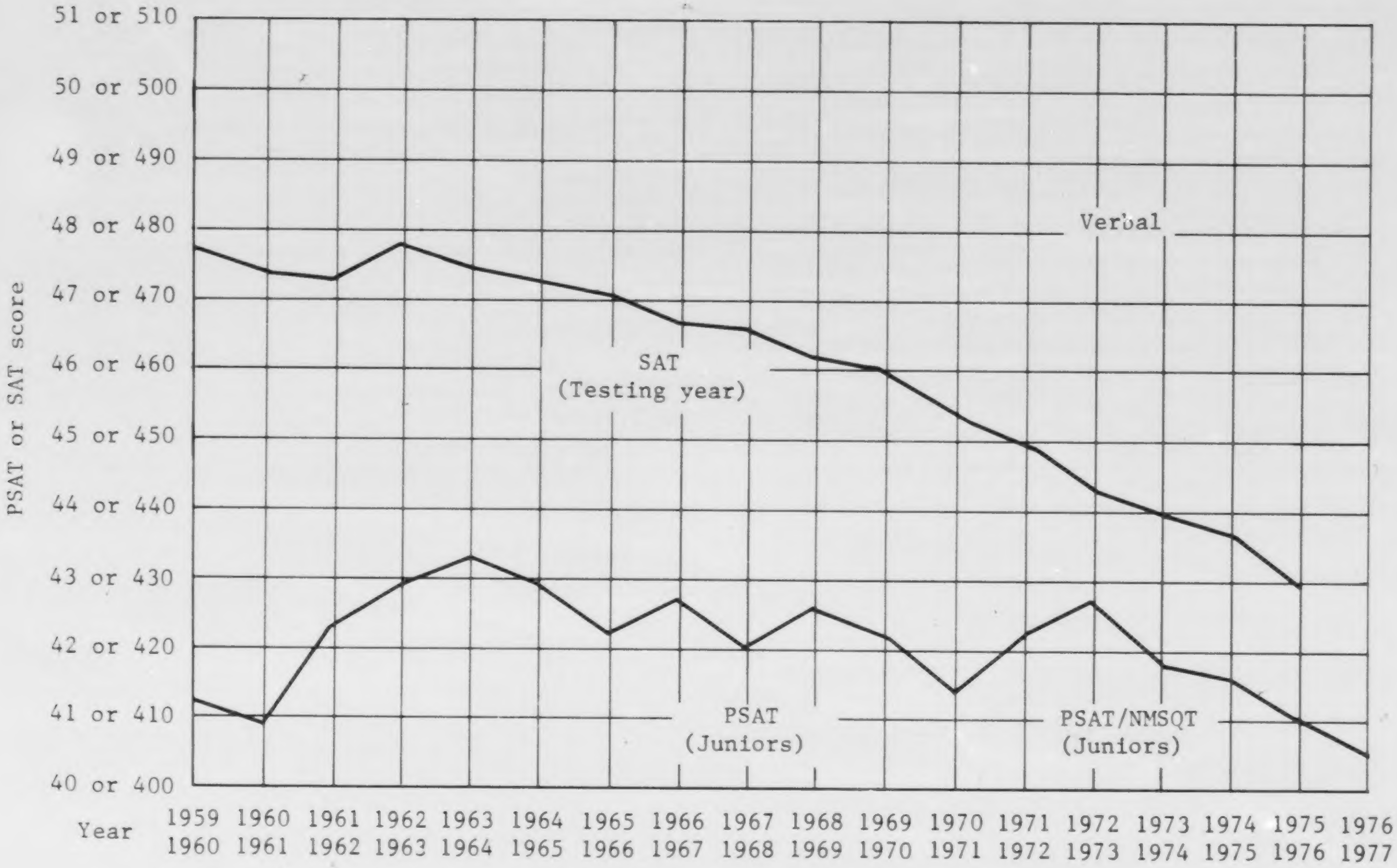
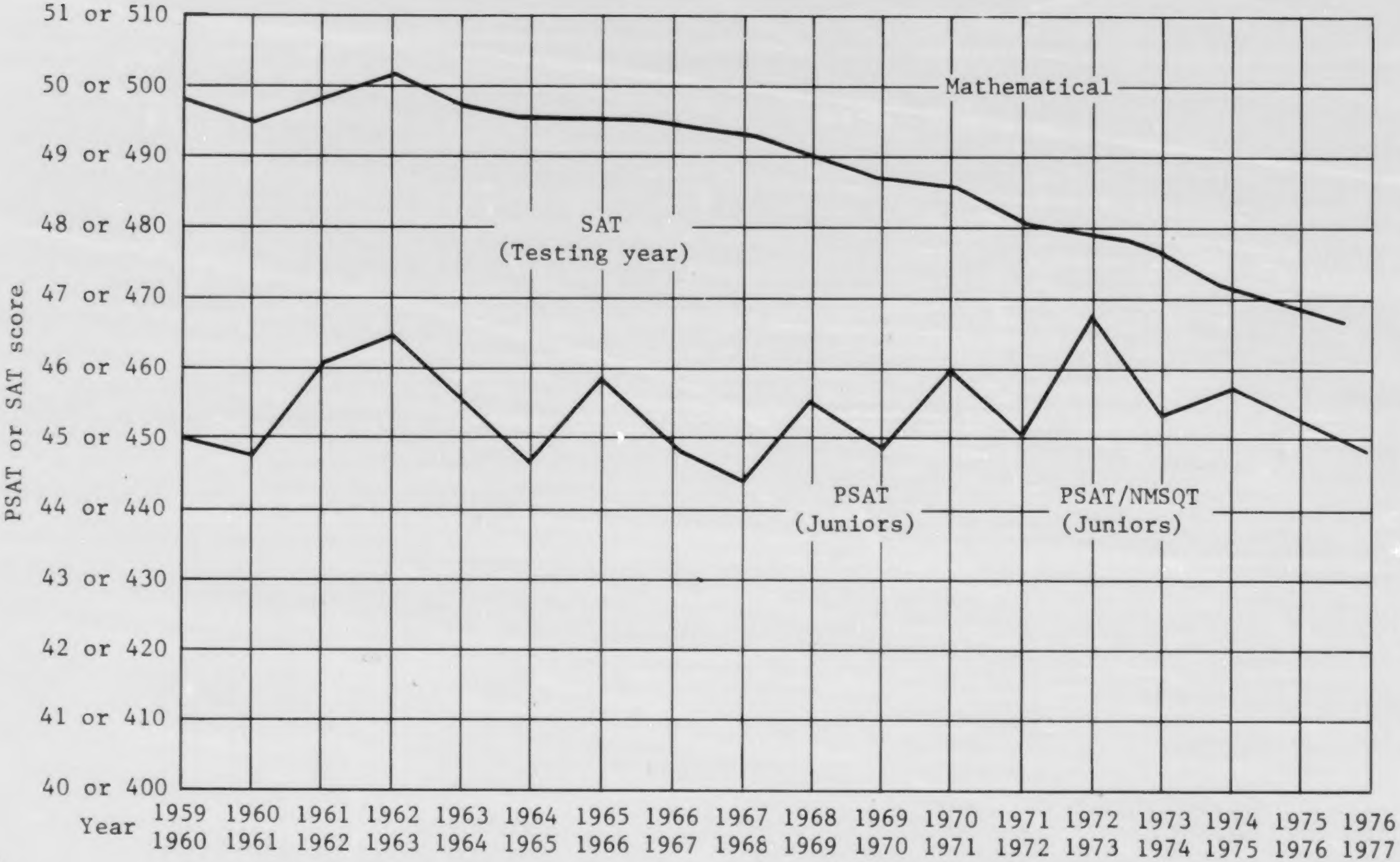

51 or 510
50 or 500
49 or 490
48 or 480
47 or 470
46 or 460
45 or 450
44 or 440
43 or 430
42 or 420
41 or 410
40 or 400
PSAT or SAT score
Verbal
SAT
(Testing year)
PSAT
(Juniors)
PSAT/NMSQT
(Juniors)
Year
1959 1960
1960 1961
1961 1962
1962 1963
1963 1964
1964 1965
1965 1966
1966 1967
1967 1968
1968 1969
1969 1970
1970 1971
1971 1972
1972 1973
1973 1974
1974 1975
1975 1976
1976 1977

Figure 2


51 or 510
50 or 500
49 or 490
48 or 480
47 or 470
46 or 460
45 or 450
44 or 440
43 or 430
42 or 420
41 or 410
40 or 400
PSAT or SAT score
Mathematical
SAT
(Testing year)
PSAT
(Juniors)
PSAT/NMSQT
(Juniors)
Year
1959 1960
1960 1961
1961 1962
1962 1963
1963 1964
1964 1965
1965 1966
1966 1967
1967 1968
1968 1969
1969 1970
1970 1971
1971 1972
1972 1973
1973 1974
1974 1975
1975 1976
1976 1977

A Summary of SAT Score Statistics for College Board Candidates

---

Rex Jackson
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

October 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

Introduction

This report summarizes certain available data concerning performance of students taking the Scholastic Aptitude Test (SAT) with an emphasis on the years from the mid-1960s to the present - the years during which the mean scores of students taking the SAT have declined. Section 1 gives an overview of trends in SAT performance from 1956-57 to the present. Section 2 presents certain additional comparisons of SAT candidates over the years 1971-72 to 1975-76 that were made possible by the introduction of the Student Descriptive Questionnaire and the Summary Reporting Service in 1971-72. Section 3 presents comparisons of SAT performance of students from the 1972-73, 1973-74, 1974-75, and 1975-76 high school classes grouped according to various biographical characteristics.

Section 1. Overview

Table 1 presents the means, standard deviations, and numbers of cases for all candidates taking the SAT, by year, from 1956-57 through 1975-76. The term "case" is used here to denote an attendance at an SAT administration. Thus, if a candidate attended two administrations within a year, he (she) is counted as two cases. In addition, candidates taking the test in more than one year (e.g., students taking the test as juniors and repeating as seniors) will appear in the counts for each of the years.

Although statistics based on testing year groups are not the most satisfactory ones, they are presented here because they are the ones for which the longest term comparisons can be made using available data. Tables presented later in this report will give shorter term comparisons for students grouped in a more meaningful way.

The following trends are apparent from Table 1:

a. A dramatic rate of increase in number of students taking the SAT from 1956-57 to 1964-65 and a slower rate of increase from 1964-65 to 1969-70; then a decline in numbers of candidates to 1973-74.

b. Apparently stable mean SAT verbal and mathematical scores for testing year candidates from 1956-57 to about 1962-63; then a fairly steady decline in means for both scores through 1975-76.

c. Somewhat larger declines for verbal than for mathematical mean scores, and somewhat larger declines for both means in more recent years.

It is worthy of note, and somewhat surprising, that mean scores remained relatively stable during the period 1956-57 to 1962-63 when numbers of students taking the test were increasing significantly from year to year, since under these conditions the groups of students taking the test would be expected to differ from year to year on a number of characteristics that might be related to test performance. Paradoxically, the decline in mean scores occurred mainly in a period when the numbers of students taking the test were more nearly constant from year to year.

Tables A.1 and A.2 in the appendix present more detailed breakdowns for the testing years from 1966-67 through 1975-76, giving information comparable to that in Table 1 for students classified by sex and grade level. Table 2 presents frequency and percentage distributions of

Table 1. Means, Standard Deviations, and Numbers of Cases*
for all Candidates Taking the SAT, by Year, 1956-57 through 1975-76

Testing year	SAT-verbal		SAT-mathematical		
	Mean	SD	Mean	SD	N/1000
1956-57	473	105	496	111	270.5
1957-58	472	106	496	109	376.8
1958-59	475	106	498	110	469.7
1959-60	477	108	498	109	564.2
1960-61	474	108	495	109	716.3
1961-62	473	109	498	112	802.5
1962-63	478	109	502	111	933.1
1963-64	475	108	498	113	1,163.9
1964-65	473	108	496	112	1,361.2
1965-66	471	108	496	111	1,381.4
1966-67	467	109	495	110	1,422.5
1967-68	466	108	494	112	1,543.8
1968-69	462	108	491	111	1,585.6
1969-70	460	110	488	113	1,605.9
1970-71	454	109	487	113	1,537.2
1971-72	450	108	482	112	1,459.9
1972-73	443	107	481	112	1,398.4
1973-74	440	107	478	115	1,354.0
1974-75	437	108	473	114	1,371.2
1975-76	429	109	470	118	1,415.0

*The term "case" denotes attendance at an SAT administration. Thus, if an individual was tested twice, he (she) is counted twice in these tabulations.

verbal and mathematical scores for the period 1966-67 through 1975-76. These distributions (also based on all candidates tested in a given year) indicate that the declines in means are not solely attributable to increased numbers of lower scoring students. On the verbal test, the percentage of students scoring over 600 declined from over 12 percent in 1966-67 to under 8 percent in 1975-76 -- a decline in absolute numbers from nearly 180,000 to less than 110,000. On the mathematical test, the comparable decline was smaller with about 18 percent scoring over 600 in 1966-67 compared to about 15 percent in 1975-76.

It should be noted that although there has been a study decline over this period in percentage of students scoring over 600, the sharp declines in numbers of students scoring over 600 began in 1970-71 for the verbal test and 1971-72 for the mathematical test. Prior to 1970-71, the total number of candidates was increasing from year to year, the increase being evident mainly in lower score intervals. The numbers of students scoring over 600 were relatively constant, but these students were a smaller percentage of the total test-taking group.

As was indicated earlier, statistics for testing-year groups are somewhat unsatisfactory. Because of this, since 1971-72, SAT scores have also been analyzed for students grouped by high school graduating classes (or "cohorts"). When data are presented in this way, the figures for any given year apply to all students from the given high school classes who took the SAT at any time in their high school careers (through the spring of the senior year). For example, statistics for the 1975-76 class are based on students taking the test as seniors in 1975-76, as juniors in 1974-75, and (in a very small number of cases) as sophomores in 1973-74. Each student is counted only once, and the most recent score is used for students repeating the test.

Although statistics for cohort groups have been calculated only since 1971-72, it was possible to use information on students repeating the test to estimate the comparable statistics for graduating classes from 1966-67 through 1970-71. Because these estimates were formed by actually subtracting out the known scores for repeaters, they are judged to be very close estimates of the true figures, and, in fact, comparison of the results of this estimation procedure with the actual results for later years suggests that the means are probably within one point of the true values.

Table 3 presents cohort group means, standard deviations and number of cases for SAT-V and SAT-M for the period 1966-67 through 1975-76. Figures are given separately for males, females and total groups. The total group means for the high school classes are very similar to those for the testing year groups given earlier and indicate a decline of 35 points in the verbal mean and of 20 points in the mathematical mean since 1966-67. Although means for cohort groups are not available for earlier years, the close correspondence of means for cohort and testing year groups for the years for which both are available suggests that the total decline in high school class means from 1962-63 to the present would be very close to the 49 points on SAT-V and 32 points on SAT-M observed for candidates grouped on a testing year basis.

For the verbal test, the decline has been somewhat greater for females than for males -- 38 compared to 30 points for high school classes from 1966-67 through 1975-76. For the mathematical test, the declines in mean scores for the two sexes are very similar, being slightly larger for females. It should be noted that the numbers of females relative to numbers of males taking the SAT have been increasing over this period. Girls taking the SAT in the mid-1960s may well have been a more select group relative to boys taking the SAT than the comparable more recent groups.

Section 2. Five-Year Comparisons of SAT Candidates

The introduction of a Student Descriptive Questionnaire (SDQ) in 1971-72 and a Summary Reporting Service permits a number of additional comparisons of students from the classes of 1971-72 through 1975-76. Table 4 presents a comparison of responses to a number of SDQ questions for students from those classes who completed the SDQ. Two particular cautions are in order in interpreting these data:

a. SDQ results for the 1971-72 class are based on students participating as seniors and do not include more than 200,000 students from that class who participated as juniors only in 1970-71. Thus the figures for this year are not strictly comparable to those for later years.

b. Not all students who take the SAT complete the SDQ, and not all SDQ

Table 2. Numbers (in thousands) and Percentages of Candidates in Given Score Ranges on the Scholastic Aptitude Test, 1966-67 through 1975-76

a. Verbal

Score range	1966-67		1967-68		1968-69		1969-70		1970-71	
	N	(percent)	N	(percent)	N	(percent)	N	(percent)	N	(percent)
700-800. . .	33.2	(2.3)	30.5	(2.0)	27.9	(1.8)	27.6	(1.7)	25.6	(1.7)
600-699. . .	145.3	(10.2)	156.7	(10.1)	157.0	(9.9)	161.7	(10.1)	145.2	(9.4)
500-599. . .	339.4	(23.9)	360.8	(23.4)	379.3	(23.9)	378.4	(23.6)	361.8	(23.5)
400-499. . .	503.5	(35.4)	534.9	(34.6)	548.8	(34.6)	530.9	(33.1)	520.2	(33.8)
300-399. . .	325.4	(22.9)	378.1	(24.5)	377.7	(23.8)	401.7	(25.0)	385.7	(25.1)
200-299. . .	75.7	(5.3)	82.9	(5.4)	94.9	(6.0)	105.4	(6.6)	98.6	(6.4)
	1422.5		1543.8		1585.6		1605.9		1537.2	

Score range	1971-72		1972-73		1973-74		1974-75		1975-76	
	N	(percent)	N	(percent)	N	(percent)	N	(percent)	N	(percent)
700-800. . .	20.0	(1.4)	16.0	(1.1)	16.8	(1.2)	14.8	(1.1)	15.9	(1.1)
600-699. . .	130.7	(9.0)	112.2	(8.0)	104.3	(7.7)	95.5	(7.0)	92.3	(6.5)
500-599. . .	332.5	(22.8)	300.1	(21.5)	279.6	(20.6)	289.4	(21.1)	273.1	(19.3)
400-499. . .	493.4	(33.8)	470.2	(33.6)	463.7	(34.2)	468.5	(34.2)	474.2	(33.5)
300-399. . .	378.6	(25.9)	390.7	(27.9)	378.5	(28.0)	376.4	(27.4)	405.8	(28.7)
200-299. . .	104.7	(7.2)	109.2	(7.8)	111.2	(8.2)	126.7	(9.2)	153.6	(10.9
	1459.9		1398.4		1354.0		1371.2		1415.0	

(continued)

Table 2. (continued)

b. Mathematical

Score range	1966-67 N	(percent)	1967-68 N	(percent)	1968-69 N	(percent)	1969-70 N	(percent)	1970-71 N	(percent)
700-800. . .	55.5	(3.9)	56.2	(3.6)	55.9	(3.5)	56.0	(3.5)	55.7	(3.6)
600-699. . .	199.0	(14.0)	225.9	(14.6)	224.0	(14.1)	224.4	(14.0)	222.9	(14.5)
500-599. . .	413.3	(29.1)	442.2	(28.6)	454.1	(28.6)	457.7	(28.5)	428.3	(27.9)
400-499. . .	465.8	(32.7)	478.5	(31.0)	495.1	(31.2)	497.6	(31.0)	483.2	(31.4)
300-399. . .	248.1	(17.4)	282.4	(18.3)	296.3	(18.7)	303.5	(18.9)	291.9	(19.0)
200-299. . .	40.7	(2.9)	58.6	(3.8)	60.2	(3.8)	66.6	(4.1)	55.2	(3.6)
	1422.5		1543.8		1585.6		1605.9		1537.2	

Score range	1971-72 N	(percent)	1972-73 N	(percent)	1973-74 N	(percent)	1974-75 N	(percent)	1975-76 N	(percent)
700-800. . .	44.3	(3.0)	44.8	(3.2)	46.3	(3.4)	45.5	(3.3)	44.1	(3.1)
600-699. . .	203.2	(13.9)	184.0	(13.2)	175.9	(13.0)	168.7	(12.3)	169.8	(12.0)
500-599. . .	415.3	(28.4)	402.8	(28.8)	365.2	(27.0)	360.6	(26.3)	371.2	(26.2)
400-499. . .	449.4	(30.8)	429.1	(30.7)	413.1	(30.5)	413.1	(30.1)	425.6	(30.1)
300-399. . .	290.2	(19.9)	280.8	(20.1)	291.6	(21.5)	315.1	(23.0)	314.3	(22.2)
200-299. . .	57.6	(3.9)	57.0	(4.1)	61.9	(4.6)	68.1	(5.0)	89.9	(6.4)
	1459.9		1398.4		1354.0		1371.2		1415.0	

Table 3. Means, Standard Deviations, and Numbers of Cases, SAT-Verbal and SAT-Mathematical, for High School Classes, 1966-67 through 1975-76

SAT-verbal

	Males			Females			Total		
Class	Mean	SD	N/1000	Mean	SD	N/1000	Mean	SD	N/1000
1966-67. . . .	463	109		468	108		466	108	
1967-68. . . .	464	111		466	108		466	110	
1968-69. . . .	459	110		466	109		463	110	
1969-70. . . .	459	111		461	110		460	111	
1970-71. . . .	454	111		457	109		455	111	
1971-72. . . .	454	112	523.8	452	110	499.1	453	111	1022.8
1972-73. . . .	446	110	515.9	443	107	498.9	445	108	1014.9
1973-74. . . .	447	111	492.6	442	108	492.6	444	110	985.2
1974-75. . . .	437	111	496.9	431	108	499.6	434	109	996.4
1975-76. . . .	433	111	494.6	430	110	505.2	431	110	999.8

SAT-mathematical

	Males			Females			Total		
Class	Mean	SD	N/1000	Mean	SD	N/1000	Mean	SD	N/1000
1966-67. . . .	514	113		467	103		492	111	
1967-68. . . .	512	116		470	107		492	114	
1968-69. . . .	513	116		470	105		493	113	
1969-70. . . .	509	117		465	106		488	114	
1970-71. . . .	507	117		466	105		488	113	
1971-72. . . .	505	118	523.7	461	108	499.0	484	11[illegible]	1022.7
1972-73. . . .	502	117	515.9	460	106	498.8	481	113	1014.7
1973-74. . . .	501	119	492.6	459	108	492.5	480	116	985.1
1974-75. . . .	495	119	496.8	449	107	499.5	472	115	996.4
1975-76. . . .	497	123	494.6	446	111	505.2	472	120	999.8

All figures for 1966-67 through 1970-71 are estimates. See text.

respondents take the SAT. Thus, summaries of SDQ responses for any year are not based on a group identical to the SAT candidates from that class. However, the overlap of the two groups is very large, and summaries of SDQ responses may be taken as reasonably accurate indicators of the composition of the SAT candidate group.

It appears from the self-report information in Table 4 that SDQ respondent (and probably SAT candidates) from the five classes are similar in terms of ethnic background, rank in class, and educational goals. There is, however, a trend away from liberal arts to other fields -- primarily career-oriented fields such as business, education, and health. (A change in the choices for this question starting with the 1974-75 SDQ may have had some effect on responses, but the consistency of the trend over time is marked.) There is also a rise in reported parental income as would be expected because of inflation.

Table 4. Summary of Responses to Selected Questions on the Student Descriptive Questionnaire for Respondents from the Classes of 1971-72 through 1975-76

	1971-72[1]	1972-73	1973-74	1974-75	1975-76
1. Ethnic background:					
Percent minority. .	13	13	14	14	15
2. Rank in class:					
Percent in top fifth	48	47	48	48	48
3. Educational goals:					
Percent two year. .	8	7	7	7	6
Percent four year .	30	29	29	29	29
Percent graduate/ professional. . .	37	36	39	40	41
Percent undecided .	25	27	26	24	24
4. Intended field of study[2]:					
Percent liberal arts.	49	44	44	38	33
Percent engineering . . .	6	6	5	7	8
Percent other fields.	40	45	43	49	53
Percent undecided .	5	7	9	7	5
5. Parental income:					
Average	NA	\$15,900	\$17,600	\$19,000	\$20,000

Note: All percentages are based on number responding. Percentages may total to more or less than 100 for items 3 and 4 because of rounding.

1. 1971-72 figures based only on responses of those members of the class participating as seniors.

2. Calculated by collapsing more detailed categories reported in College-Bound Seniors for each year.

With the introduction of the Summary Reporting Service summaries of SAT performance for students grouped by state and region have also become available. The regional summaries for students from the classes of 1971-72 through 1975-76 are given in Table 5. It is apparent that the decline in mean scores was not limited to any one section of the country with all regions showing some decline for both SAT-V and SAT-M over the four years. The declines for New England, the Middle States, the South, and the Southwest are very close to the national figures. Those for the West are larger, while those for the Midwest are smaller.

Table 5. Means, Standard Deviations, and Numbers of Cases of SAT-V and SAT-M Scores for High School Classes, 1971-72 through 1975-76, by Region

		SAT-V		SAT-M		
	Class	Mean	SD	Mean	SD	N/1000
New England.	1971-72	456	108	482	111	115.1
	1972-73	447	105	482	109	112.2
	1973-74	447	107	479	112	111.3
	1974-75	437	105	471	110	115.7
	1975-76	435	107	472	116	117.2
Middle States.	1971-72	453	110	484	113	358.7
	1972-73	447	107	484	112	341.6
	1973-74	445	108	481	115	335.1
	1974-75	433	107	471	113	342.8
	1975-76	431	108	472	118	341.8
South.	1971-72	429	114	455	116	159.9
	1972-73	424	111	455	113	153.6
	1973-74	426	112	455	115	151.5
	1974-75	416	111	447	114	159.6
	1975-76	413	113	445	119	165.7
Midwest.	1971-72	464	108	500	116	193.2
	1972-73	458	106	498	114	177.9
	1973-74	459	108	502	117	154.5
	1974-75	451	108	497	118	142.8
	1975-76	451	109	499	122	140.7
Southwest.	1971-72	452	111	483	112	55.9
	1972-73	443	108	479	111	56.8
	1973-74	444	109	480	115	53.1
	1974-75	436	109	472	114	54.2
	1975-76	433	110	471	117	54.8
West	1971-72	467	110	497	116	131.8
	1972-73	455	110	490	115	135.5
	1973-74	454	111	488	116	136.1
	1974-75	441	110	478	117	144.5
	1975-76	435	112	478	120	147.8

Section 3. Comparisons of Students from the 1972-73, 1973-74, 1974-75, and 1975-76 Classes Grouped by SDQ Responses

Table 6 gives SAT verbal and mathematical means and percentages of cases for students from the 1972-73, 1973-74 1974-75, and 1975-76 classes grouped according to SDQ responses indicating high school type, type of high school program, size of high school class, rank in class, weighted grade average, number of scholastic honors earned, educational goals, intended college major, and parental income. These data are based on samples of nearly 100,000 cases each drawn from the nearly 1,000,000 students from each of these classes taking the SAT. Comparison of sample characteristics with known characteristics of the total groups for each year indicates that the samples very closely represent the total groups. In some cases where the questionnaires differed, categories have been combined to yield comparable groups. In addition, choices selected by very few students have, in some cases, been grouped together.

Table 6 also shows the differences in SAT verbal and mathematical means between 1972-73 and 1975-76 for each subgroup. In most cases, the declines for subgroups are not very different from the 14 point mean verbal decline and 9 point mean mathematical decline for the total groups over this period. All but a few of the small subgroups show some decline for both scores.

The general picture presented by Table 6 is of a highly consistent pattern of declining scores across nearly all subgroups identified by these SDQ responses. Some detailed observations are warranted, however:

a. Differences in response rates and in the score levels of non-respondents across the years appear to have had an effect on some year-to-year comparisons and should be taken into account when the data are examined in detail.

b. The results for students grouped by self-reported class rank and high school average suggest that groups of students reporting relatively low academic standing may have somewhat larger declines. For at least high school average, however, the possible effects of grade inflation should be considered. If it is assumed that inflation has occurred, any correction for such inflation would have the effect of moving more of the recent students to lower grade intervals. This would result in somewhat higher within group mean scores for the more recent year and thus reduce the size of the apparent decline.

Another interesting characteristic of the tables for class rank and high school average is a tendency for the lower achieving groups to exhibit mathematical declines that equal or exceed verbal declines.

c. The trend toward career fields and away from arts and sciences when students report their intended college major is apparent here as it was in Table 4. In addition, it can be seen that the within group declines are relatively small when students are grouped according to whether their intended major is a career or an arts and sciences field. These facts taken together are consistent with a theory attributing part of the decline in scores to changing composition of the test-taking population. That is, one interpretation is that career-oriented college-bound students and liberal arts students each have small mean score declines over these four years and that the larger total group declines are due to the entry of considerable numbers of generally lower scoring career-oriented

Table 6. SAT Mean Scores for Samples of Students from the 1972-73, 1973-74, 1974-75, and 1975-76 Classes, Classified by Responses to Selected SDQ Questions

	1972-73			1973-74			1974-75			1975-76			Differences in means 1972-73 to 1975-76	
	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	SAT-V	SAT-M
1. High school type:														
Public	63.0	449	488	67.7	446	484	65.9	437	477	63.5	437	480	12	8
Private	13.4	460	488	14.6	458	484	14.5	448	476	13.6	448	479	12	9
No response	23.6	425	460	17.7	426	461	19.6	416	451	22.9	407	445	18	15
2. High school program:														
Academic	61.6	464	503	64.3	463	500	62.8	454	493	59.6	455	499	9	4
General	9.9	403	436	11.9	400	433	11.6	392	427	11.5	389	426	14	10
Career/other	3.8	372	399	4.9	375	401	5.2	366	396	5.4	365	393	7	6
No response	24.8	424	459	18.8	425	461	20.4	415	450	23.5	407	444	17	15
3. Class size:														
Less than 100	6.8	448	479	7.3	446	478	7.0	435	467	6.9	432	466	16	13
100-249	15.9	449	484	15.9	445	478	15.8	435	472	15.4	436	475	13	9
250-499	22.1	455	493	23.4	451	488	23.6	443	482	22.9	443	485	12	8
500-749	15.1	457	497	16.0	457	494	16.2	444	485	15.9	445	490	12	7
750 or more	14.7	443	483	17.6	441	479	16.2	433	473	14.8	431	474	12	9
No response	25.5	426	460	19.9	427	461	21.1	417	452	24.1	409	446	17	14
4. Class rank:														
Top tenth	15.3	538	584	16.5	538	585	16.9	524	575	17.0	524	580	14	4
Second tenth	15.9	473	512	18.6	466	505	20.2	455	496	19.1	455	500	18	12
Second fifth	17.2	435	473	20.7	424	457	22.7	414	449	20.7	414	452	21	21
Third fifth	14.2	391	423	15.3	382	409	15.6	374	402	15.8	371	400	20	23
Bottom two-fifths	2.1	370	399	1.7	360	385	1.6	350	376	1.4	347	372	23	27
No response	35.4	422	455	27.2	428	462	23.0	417	451	26.0	408	445	14	10

(continued)

Table 6 (Continued)

	1972-73			1973-74			1974-75			1975-76			Differences in means 1972-73 to 1975-76	
	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	SAT-V	SAT-M
5. Average grades:														
3.75-4.00	11.5	539	580	12.6	541	585	12.7	528	577	13.1	527	580	12	0
3.50-3.74	9.2	496	533	9.9	496	533	10.2	483	522	10.5	482	525	14	8
3.25-3.49	9.7	472	512	10.6	469	507	10.9	459	499	10.8	455	499	17	13
3.00-3.24	12.7	447	486	14.0	442	478	14.1	430	468	13.5	425	467	22	19
2.75-2.99	8.6	428	463	9.5	421	452	9.6	407	441	9.0	406	442	22	21
2.50-2.74	9.0	406	442	9.6	401	432	9.2	391	423	8.4	388	421	18	21
2.25-2.49	5.9	391	428	6.1	383	417	5.8	375	408	5.0	368	403	23	25
2.00-2.24	4.5	377	410	4.5	368	400	4.0	359	394	3.6	356	388	21	22
Less than 2.00	3.2	364	397	3.1	356	388	2.5	347	378	2.1	340	373	24	24
No response	25.7	425	459	20.1	427	461	21.0	416	451	24.0	408	445	17	14
6. Scholastic honors:														
0	37.3	423	460	38.0	418	454	37.3	411	448	34.8	410	449	13	11
1-2	22.1	465	503	24.0	461	496	24.6	448	488	23.9	448	491	17	12
3-4	8.1	495	531	9.1	489	526	9.4	476	516	9.4	476	520	19	11
5-6	3.0	512	549	3.3	503	540	3.4	491	532	3.5	492	535	20	14
7 or more	2.7	524	562	3.2	518	557	3.2	507	544	3.4	506	548	18	14
No response	26.8	427	462	22.3	432	467	22.1	420	455	25.0	411	448	16	14
7. Educational goals:														
2-year technical	3.8	382	414	3.6	376	409	3.4	367	405	3.3	366	402	16	12
2-year liberal arts	1.4	382	405	1.6	377	395	1.5	369	392	1.4	370	388	12	17
Bachelors degree	22.0	440	476	23.1	436	471	23.1	425	462	21.9	424	465	16	11
Masters	14.2	469	506	15.4	467	502	16.5	454	493	16.5	454	497	15	9
Doctorate/professional degree	12.6	499	540	14.5	498	537	14.9	486	529	14.5	486	533	13	7
Other/undecided	20.1	439	477	20.8	433	469	19.1	424	462	17.7	423	462	16	15
No response	26.0	425	460	21.1	427	462	21.7	417	452	24.7	408	446	17	14

(continued)

Table 6 (Continued)

	1972-73			1973-74			1974-75			1975-76			Differences in means 1972-73 to 1975-76	
	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	SAT-V	SAT-M
8. Intended college major[1]:														
Agriculture.	1.7	427	471	1.9	426	468	1.9	423	459	1.9	426	462	1	9
Architecture	1.4	438	515	1.5	436	504	1.5	430	507	1.5	434	517	4	(+2)
Business	7.1	409	463	8.2	406	457	8.1	406	461	8.0	413	471	(+4)	(+8)
Education.	7.5	418	449	6.5	417	448	7.1	405	434	6.6	405	431	13	18
Home economics	1.2	413	441	1.3	407	437	0.7	409	442	0.6	410	436	3	5
Journalism/communication	1.6	476	483	1.8	472	470	2.2	458	461	2.2	465	468	11	15
Nursing/health	6.8	419	444	8.2	417	445	8.7	410	444	9.1	409	444	10	0
Vocational tech. . . .	1.9	400	450	1.9	389	436	1.9	370	405	2.0	365	392	35	58
Miscellaneous career .	.0	-	-	0.7	444	520	4.1	419	460	6.1	419	463	na	na
Subtotal career	29.2	419	457	32.0	417	455	36.4	411	450	37.9	414	453	5	4
Art.	2.8	440	451	2.8	432	441	1.8	435	445	1.2	438	443	2	8
Biological science . .	9.2	493	533	10.1	488	528	8.8	481	525	7.6	483	530	10	3
English.	2.5	500	481	2.6	496	473	2.4	488	465	2.1	489	469	11	12
Foreign languages. . .	1.3	491	498	1.3	483	494	1.0	481	486	0.8	485	487	6	11
Mathematics.	2.4	481	595	2.6	479	584	1.9	463	580	1.4	472	594	9	1
Music.	2.0	465	487	2.1	462	479	1.8	448	464	1.6	452	468	13	19
Philosophy/religion. .	0.7	479	500	0.5	476	489	0.6	469	484	0.4	478	495	1	5
Physical science . . .	2.8	505	570	2.9	503	568	2.3	501	565	2.1	506	572	(+1)	(+2)
Social science	11.1	476	490	11.8	463	477	9.4	465	476	7.6	464	475	12	15
Subtotal arts and science	34.7	482	512	36.6	475	504	30.0	472	501	25.0	475	505	7	7
Engineering.	4.0	460	548	4.0	465	554	5.4	450	541	6.2	452	550	8	(+2)
Other/undecided. . . .	5.0	446	489	7.0	435	473	6.1	438	477	5.2	446	485	0	4
No response.	27.1	422	457	20.4	430	463	22.2	416	451	25.8	408	446	14	11

1. Differences between this table and the percentages reported for the major field groups in Table 4 are due to (a) the fact that samples rather than entire populations were used for this table, (b) assignment of specific fields to different categories in a small number of cases, and (3) treatment of nonrespondents as a separate category in this table. Both this table and Table 4 show clearly the trend from liberal arts to career fields, although the precise percentages may differ somewhat (even after the differences in treatment of nonrespondents is accounted for). Results in this table also differ slightly from those reported in the first version of this paper (for 1973-74 and 1974-75) because of reassignment of specific fields to new categories.

(continued)

Table 6 (Continued)

	1972-73			1973-74 .			1974-75			1975-76			Differences in means 1972-73 to 1975-76	
	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	Percent	Mean SAT-V	Mean SAT-M	SAT-V	SAT-M
9. Parental income:														
Less than $3,000 . . .	1.7	370	405	1.5	367	401	1.4	360	394	1.4	360	397	10	8
$3,000 - 5,999 . . .	4.4	402	434	3.9	391	425	3.5	383	418	3.2	378	412	24	22
6,000 - 7,499 . . .	4.2	421	456	3.6	413	445	3.1	396	431	2.8	391	425	30	31
7,500 - 8,999 . . .	5.0	433	470	4.0	425	459	3.5	411	447	3.1	404	439	29	31
9,000 - 10,499 . . .	7.1	442	479	5.9	434	470	5.0	419	454	4.0	416	452	26	27
10,500 - 11,999 . . .	6.2	447	487	5.5	440	475	5.0	427	467	4.2	426	467	21	20
12,000 - 13,499 . . .	5.8	452	491	6.0	443	478	5.8	431	472	5.0	430	472	22	19
13,500 - 14,999 . . .	5.3	461	500	4.6	450	488	4.8	439	479	4.3	436	479	25	21
15,000 - 17,999 . . .	7.5	469	507	8.2	454	492	8.9	441	482	8.2	440	482	29	25
18,000 or more	17.0	483	523	22.5	467	503	26.9	463	502	28.9	462	507	21	16
No response.	36.0	431	465	34.3	444	479	32.1	427	463	34.9	420	457	11	8

students into the pool. It would also appear to be possible, however, for these results to have been produced simply by changing preferences among college-bound students. The groups taking the SAT may have changed little demographically, but more of the middle scoring students may be expressing interest in career fields compared to arts and sciences, resulting in larger numbers in the career group and smaller within group declines for both groups. It may be noted that this second explanation does not require the loss of the traditional college-bound student from the pool (but merely a change in interests), while the first explanation suggesting the entry of a new group would entail a large loss in traditional arts and sciences students since the total N's have been fairly constant. The data reported here do not permit a certain choice between these possibilities. In addition, the question is complicated by a change in this portion of the questionnaire in 1974-75 which may have affected patterns of response. This issue will be explored in greater depth in a separate paper.

d. The relatively large declines for many of the intervals for parental income would be moderated somewhat if inflation were taken into account.

Similar analyses were carried out for a number of additional SDQ variables, but are not reported here. Included among these other variables were reported grades in specific courses and number of courses taken in certain subject matter areas, participation in extracurricular activities, and other SDQ responses. In all cases for which the numbers of students responding were sufficient to permit accurate estimation of mean differences and for which questionnaires were consistent over the years, the means for the more recent subgroups were lower than were those for the earlier subgroups.

Summary

1. SAT verbal and mathematical mean scores have declined steadily for groups of candidates taking the test since 1962-63. When data are analyzed for members of high school classes taking the SAT (rather than all candidates tested in a given year) similar declines are observed.
2. Examination of score distributions indicates that the declines are not attributable solely to increased numbers of low scoring students. Both percentages and numbers of high scoring students have declined -- with the decline in numbers beginning in about 1970-71.
3. Responses of students from the last five high school classes to the Student Descriptive Questionnaire suggest that the students from these classes taking the SAT are similar in terms of several background characteristics, but there is a marked trend toward career fields and away from arts and sciences in choice of an intended major field. Whether this last finding is simply due to changes in patterns of choice among the customary SAT taking group or to entry of new kinds of students into the pool (or to a combination of both) is uncertain.
4. Mean scores of students from the last four high school classes taking the SAT have declined in all regions of the country though there are some regional differences in amount of decline.
5. When samples of students from the last four high school classes are grouped according to responses to a number of questions bearing on their backgrounds, high school preparation, and educational plans, nearly all subgroups show some decline from 1972-73 to 1975-76, with the declines in most cases close to those for the total group.

APPENDIX

Supplementary Tables

Table A1. Ten-Year Summary Statistics for SAT-verbal, All Candidates Tested 1966-67 through 1975-76

Year and grade level	Male			Female			Total[a]		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
1966-67:									
Juniors	262,669	473	108	235,963	479	106	498,636	476	108
Seniors	475,148	461	109	384,828	466	109	859,982	463	109
All grades[b]	777,250	465	109	645,217	471	108	1,422,479	467	109
1967-68:									
Juniors	307,493	468	108	282,382	473	103	589,879	470	105
Seniors	490,737	463	111	399,977	464	109	890,721	463	110
All grades[b]	837,315	464	110	706,511	467	106	1,543,839	466	108
1968-69:									
Juniors	314,223	465	106	293,667	469	105	607,896	467	105
Seniors	498,462	457	110	415,925	464	110	914,395	460	110
All grades[b]	852,105	460	108	733,439	466	108	1,585,560	462	108
1969-70:									
Juniors	310,715	464	109	297,908	466	105	608,629	465	107
Seniors	504,588	457	112	429,712	458	111	934,324	458	112
All grades[b]	854,275	459	111	751,539	461	109	1,605,851	460	110
1970-71:									
Juniors	307,109	461	108	302,493	462	104	609,606	461	106
Seniors	460,715	449	110	403,251	452	109	863,981	450	110
All grades[b]	805,706	453	110	731,456	456	107	1,537,186	454	109
1971-72:									
Juniors	293,458	456	105	292,302	454	103	585,774	455	104
Seniors	430,208	447	111	390,534	444	110	820,763	446	110
All grades[b]	754,404	451	109	705,484	448	108	1,459,924	450	108

(continued)

Table Al. (Continued)

Year and grade level	Male			Female			Total[a]		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
1972-73:									
Juniors	280,566	450	105	285,738	448	102	570,066	449	104
Seniors	400,108	441	110	373,242	437	107	779,205	439	108
All grades[b]	708,006	445	108	679,905	442	105	1,398,354	443	107
1973-74:									
Juniors	276,934	444	104	279,294	437	101	556,602	441	103
Seniors	369,627	443	111	360,272	437	109	730,434	440	110
All grades[b]	682,870	443	109	670,044	437	106	1,353,971	440	107
1974-75:									
Juniors	274,049	446	104	274,135	448	103	548,702	447	104
Seniors	379,095	432	111	373,183	428	108	753,180	430	110
All grades[b]	690,984	437	109	678,651	436	107	1,371,232	437	108
1975-76:									
Juniors	278,654	440	106	285,665	440	105	564,790	440	105
Seniors	386,896	426	110	387,597	420	110	775,348	423	110
All grades[b]	705,519	431	109	707,977	428	109	1,414,961	429	109

a. The total N for "Total" includes candidates who did not indicate sex.

b. The total N for "All Grades" includes candidates who are not juniors or seniors.

Table A2. Ten-Year Summary Statistics for SAT-mathematical, All Candidates Tested 1966-67 through 1975-76

Year and grade level	Male			Female			Total[a]		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
1966-67:									
Juniors	262,669	525	111	235,763	485	99	498,636	506	107
Seniors	475,148	511	113	384,828	463	103	859,982	490	111
All grades[b]	777,250	515	113	645,217	470	102	1,422,479	495	110
1967-68:									
Juniors	307,493	522	110	282,382	483	100	589,879	504	107
Seniors	490,737	509	117	399,977	466	107	890,721	490	115
All grades[b]	837,315	513	115	706,511	471	105	1,543,839	494	112
1968-69:									
Juniors	314,223	515	110	293,667	474	98	607,896	495	106
Seniors	498,462	510	116	415,925	465	105	914,395	490	114
All grades[b]	852,105	511	114	733,439	468	103	1,585,560	491	111
1969-70:									
Juniors	310,715	514	112	297,908	476	100	608,629	495	108
Seniors	504,588	506	118	429,712	462	108	934,324	486	116
All grades[b]	854,275	508	116	751,539	466	105	1,605,851	488	113
1970-71:									
Juniors	307,109	513	114	302,493	474	102	609,606	494	110
Seniors	460,715	503	117	403,251	462	106	863,981	484	114
All grades[b]	805,706	505	116	731,456	466	105	1,537,186	487	113
1971-72:									
Juniors	293,458	510	110	292,302	472	100	585,774	491	107
Seniors	430,208	498	118	390,534	453	107	820,763	477	115
All grades[b]	754,404	502	115	705,484	460	105	1,459,924	482	112

(continued)

Table A2. (Continued)

Year and grade level	Male			Female			Total[a]		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
1972-73:									
Juniors	280,566	511	112	285,738	468	101	570,066	489	109
Seniors	400,108	498	117	373,242	455	106	779,205	477	114
All grades[b]	708,006	503	116	679,905	460	105	1,398,354	481	112
1973-74:									
Juniors	276,934	508	116	279,294	467	104	556,602	487	112
Seniors	369,627	495	120	360,272	454	110	730,434	474	117
All grades[b]	682,870	498	119	670,044	458	108	1,353,971	478	115
1974-75:									
Juniors	274,049	510	114	274,135	464	105	548,702	487	112
Seniors	379,095	489	118	373,183	442	106	753,180	466	114
All grades[b]	690,984	496	117	678,651	450	106	1,371,232	473	114
1975-76:									
Juniors	278,654	507	114	285,665	459	105	564,790	483	112
Seniors	386,896	490	124	387,597	438	111	775,348	464	120
All grades[b]	705,519	496	120	707,977	445	109	1,414,961	470	118

a. The total N for "Total" includes candidates who did not indicate sex.
b. The total N for "All Grades" includes candidates who are not juniors or seniors.

Trends in Mean SAT Scores for Students from the High School Classes of 1971-72 and 1975-76 Belonging to Selected School and College Groups

---

Rex Jackson
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

June 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

The summary reporting service of the College Board Admissions Testing Program permits groups of colleges to obtain group reports for Scholastic Aptitude Test (SAT) candidates from the most recent high school class who requested that their scores be reported to at least one college in the group. Similarly, high school group reports are produced based on all SAT candidates from the given class attending high schools within specified groups. These group reports supplement the individual college and school reports that are routinely produced. This paper summarizes trends in mean scores for all college and school groups for which group reports were produced in both 1971-72 and 1975-76.

Table 1 (for college groups) and Table 2 (for school groups) give the means and standard deviations of SAT scores for students in the 1971-72 and 1975-76 groups, differences in means over this period, and the ratios of numbers of students in more recent groups to numbers in corresponding groups for the earlier year. Actual numbers have been omitted in the interest of preserving the anonymity of those groups for which the number of students might provide a clue to group identity. The ratios are given as an aid in judging groups that have changed greatly in size and, perhaps consequently, in composition.

The two tables appear to require little discussion. Although there is a certain amount of variation, the results generally reinforce the finding of highly consistent declines during the early 1970s for nearly all subgroups studied.

Of the 34 college groups, one showed an increase in both SAT-V and SAT-M means, none showed an increase in SAT-V mean only, and two showed an increase in SAT-M mean only. Of 102 school groups, seven showed increases in means for both SAT-V and SAT-M, one showed an increase in SAT-V mean only, and 13 showed an increase in SAT-M mean only. With only a few exceptions, groups showing increases in mean SAT-V or SAT-M scores over this period had substantially smaller numbers of students in the more recent year.

In general, these data appear to be useful mainly in helping to demonstrate the pervasiveness of SAT score declines--particularly on the verbal test--over the last five years. No attempt has been made here to differentiate those groups showing increases from the remaining groups (except to note a tendency for groups with mean score increases to have declining numbers of students). A separate research project directed by Gary Echternacht has been examining the characteristics of individual schools with stable or increasing score trends. In contrast, this effort is concerned only with describing the score experience of those groups of schools and colleges formed for summary report purposes.

Table 1. Means and Standard Deviations for SAT Candidates from the High School Classes of 1971-72 and 1975-76 Having Score Reports Sent to Colleges in Selected Groups

			Class of 1971-72				Class of 1975-76				1972 to 1976 Mean Diff.		
			SAT-V		SAT-M		SAT-V		SAT-M				
Region	Group Description		M	SD	M	SD	M	SD	M	SD	SAT-V	SAT-M	N(76)/N(72)
Northeast/East	Public Four Year	College Group	399	105	423	108	400	101	433	109	-1	-10	0.71
Northeast/East	Public Four Year	College Group	427	98	447	98	398	96	427	102	29	20	0.92
Northeast/East	Public Four Year	College Group	482	97	515	99	456	94	503	104	26	12	1.09
Northeast/East	Private Four Year	University Group	553	113	593	112	536	111	592	112	17	1	1.16
Northeast/East	Private Four Year	College Group	545	111	544	106	529	106	533	104	16	11	1.02
Northeast/East	Public Four Year	University Group	488	98	524	97	452	96	504	105	36	20	0.97
Northeast/East	Public Four Year	University Group	413	92	454	100	403	89	454	102	10	0	1.01
Northeast/East	Public Four Year	University Group	523	109	563	107	494	104	554	107	29	9	0.94
Northeast/East	Public Four Year	College Group	442	93	466	95	416	93	448	99	26	18	0.91
Northeast/East	Public Four Year	College Group	467	98	499	103	440	94	483	104	27	16	0.91
Northeast/East	Public Four Year	College Group	446	95	470	95	417	92	450	101	29	20	0.80
Northeast/East	Private Four Year	College Group	575	109	609	107	554	110	602	109	21	7	1.19
Northeast/East	Private Four Year	College Group	570	111	565	106	554	111	559	106	16	6	1.07
Northeast/East	Public Four Year	College Group	427	88	461	90	407	87	453	98	20	8	0.83
Northeast/East	Catholic Four Year	College Group	505	109	525	105	485	107	518	110	20	7	1.21
Northeast/East	Private Four Year	College Group	517	114	520	107	505	112	513	107	12	7	0.80
Northeast/East	Public Four Year	University Group	479	112	509	114	453	108	496	117	26	13	0.99
Northeast/East	Public Four Year	College Group	431	88	458	90	410	87	445	96	21	13	0.75
Northeast/East	Public Two Year	College Group	402	86	425	91	386	86	417	95	16	8	0.93
Northeast/East	Public Four Year	University Group	480	102	516	104	455	103	499	110	25	17	0.91
Northeast/East	Public Two Year	College Group	426	101	460	106	407	96	449	109	19	11	0.71
Northeast/East	Public Four Year	University Group	476	97	509	110	453	95	496	104	21	11	0.96
Southeast	Public Four Year	University Group	421	111	453	120	417	112	461	125	4	-8	0.90
Southeast	Public Four Year	University	459	107	493	111	436	105	480	115	23	13	1.27
Southeast	Public Four Year	College/ University Group	475	116	508	117	463	114	505	122	12	3	1.06
Southeast	Public Four Year	College Group	418	104	452	108	401	105	441	112	17	11	1.04
Midwest	Private Four Year	College Group	537	109	555	107	513	108	546	111	24	9	0.63
Midwest	Private Four Year	College Group	540	109	558	110	523	108	555	112	17	3	0.65
Midwest	Public Four Year	University Group	457	104	495	111	451	104	502	115	6	-7	0.67
West/Southwest	Public Four Year	College/ University Group	453	107	483	112	422	105	465	114	31	18	1.04
West/Southwest	Public Four Year	University Group	500	113	534	117	472	112	523	117	28	11	1.11
West/Southwest	Public Four Year	College Group	448	107	483	111	434	107	477	114	14	6	0.91
West/Southwest	Private Four Year	College Group	476	116	509	120	462	115	505	123	14	4	0.92
West/Southwest	Public Two Year	College Group	415	104	445	107	394	101	431	106	21	14	1.79

Table 2. Means and Standard Deviations for SAT Candidates from the High School Classes of 1971-72 and 1975-76 Attending High Schools in Selected Groups

Region	Type	Location	Class of 1971-72 SAT-V M	SD	SAT-M M	SD	Class of 1975-76 SAT-V M	SD	SAT-M M	SD	1972 to 1976 Mean Diff. SAT-V	SAT-M	N(76)/N(72)
Northeast/East	Public	Small City	468	105	489	113	449	104	477	119	19	12	1.03
Northeast/East	Public	Small City	482	119	497	122	454	113	484	126	28	13	1.10
Northeast/East	Public	Small City	453	97	475	98	416	92	454	103	37	21	1.05
Northeast/East	Public	Suburbs	481	106	519	111	459	107	512	121	22	7	1.03
Northeast/East	Public	Suburbs	457	104	496	112	431	101	480	116	26	16	1.11
Northeast/East	Public	Small City	435	104	473	108	420	91	473	120	15	0	0.77
Northeast/East	Public	Small City	457	107	496	113	451	104	488	121	6	8	1.16
Northeast/East	Public	Suburbs	465	118	490	122	453	122	487	126	12	3	0.83
Northeast/East	Public	Small City	427	109	450	113	403	107	434	117	24	16	0.79
Northeast/East	Public	Small City	449	104	495	110	428	96	477	117	21	18	1.09
Northeast/East	Public	Large City	430	114	469	120	407	113	451	125	23	18	0.74
Northeast/East	Public	Small City	440	96	474	106	429	96	474	111	11	0	1.05
Northeast/East	Public	Small City	424	91	460	102	412	92	449	104	12	11	0.92
Northeast/East	Public	Large City	417	113	443	122	399	118	427	127	18	16	0.82
Northeast/East	Public	Middle Size City	371	110	398	115	350	107	387	111	21	11	0.80
Northeast/East	Public	Small City	393	91	415	99	376	97	397	103	17	18	0.70
Northeast/East	Public	Small City	425	102	466	109	414	104	457	110	11	9	0.79
Northeast/East	Public	Large City	354	100	375	100	332	102	352	99	22	23	0.75
Northeast/East	Public	Suburbs	499	110	562	117	473	106	533	115	26	29	1.24
Northeast/East	Public	Suburbs	492	113	519	115	463	112	504	120	29	15	1.08
Northeast/East	Public	Suburbs	460	109	490	113	437	105	480	117	23	10	1.12
Northeast/East	Public	Large City	392	113	423	124	350	108	384	116	42	39	0.94
Northeast/East	Public	Suburbs	429	111	458	114	416	112	452	120	13	6	0.88
Northeast/East	Public	Large City	423	118	459	126	390	132	426	129	33	33	0.89
Northeast/East	Public	Large City	330	88	348	84	300	79	329	78	30	19	0.98
Northeast/East	Catholic	Large City	473	103	486	102	435	101	458	111	38	28	0.90
Northeast/East	Catholic	Large City	449	103	456	103	427	101	452	108	22	4	0.93
Southeast	Public	Large City	368	110	396	112	332	103	359	102	36	37	0.74
Southeast	Public	Large City	457	121	484	128	458	124	480	126	-1	4	0.47
Southeast	Public	Large City	483	115	509	119	472	111	496	117	11	13	0.70
Southeast	Public	Large City	409	111	439	116	381	116	403	117	28	36	0.74
Southeast	Public	Middle Size City	484	108	521	111	480	109	518	115	4	3	0.99
Southeast	Public	Middle Size City	432	107	458	107	410	105	439	110	22	19	0.81
Midwest	Catholic	Large City	468	105	493	111	448	104	482	113	20	11	0.64
Midwest	Public	Suburbs	477	105	519	115	460	103	516	114	17	3	1.04
Midwest	Public	Suburbs	486	111	523	117	465	106	517	122	21	6	0.95

(continued)

Table 2 (continued)

			Class of 1971-72 SAT-V		SAT-M		Class of 1975-76 SAT-V		SAT-M		1972 to 1976 Mean Diff.		
Region	Type	Location	M	SD	M	SD	M	SD	M	SD	SAT-V	SAT-M	N(76)/N(72)
Midwest	Public	Suburbs	492	103	519	112	481	104	545	104	11	-26	0.55
Midwest	Public	Small City	471	96	519	109	462	96	526	113	9	-7	0.65
Midwest	Public	Small City	498	94	550	108	472	97	535	109	26	15	0.84
Midwest	Public	Suburbs	489	103	523	112	468	101	528	117	21	-5	0.81
Midwest	Public	Suburbs	495	102	536	115	484	98	545	116	11	-9	0.63
Midwest	Public	Suburbs	487	113	511	120	473	105	503	122	14	8	1.03
Midwest	Public	Middle Size City	436	98	456	107	407	98	445	113	29	11	0.81
Midwest	Public	Middle Size City	457	100	495	111	429	108	477	121	28	18	0.98
Midwest	Catholic	Middle Size City	461	99	490	105	446	98	482	104	15	8	0.85
Midwest	Public	Middle Size City	363	100	393	106	332	94	366	94	31	27	1.07
Midwest	Public	Middle Size City	427	99	461	108	402	97	452	114	25	9	0.87
Midwest	Public	Large City	415	112	435	114	380	111	408	115	35	27	0.50
Midwest	Catholic	Large City	448	101	475	107	393	112	440	126	55	35	0.61
Midwest	Public	Middle Size City	446	103	474	111	421	102	459	115	25	15	0.86
Midwest	Catholic	Small City	506	100	522	105	484	104	505	107	22	17	0.71
Midwest	Public	Large City	506	106	536	114	510	114	554	117	-4	-18	0.56
Midwest	Public	Large City	436	124	464	127	440	118	477	126	-4	-13	0.78
Midwest	Public	Large City	393	110	423	112	393	112	440	126	0	-17	0.52
Midwest	Public	Middle Size City	439	112	470	118	426	131	473	137	13	-3	0.53
Midwest	Public	Middle Size City	434	125	473	133	432	120	471	128	2	2	0.57
Midwest	Public	Middle Size City	420	112	458	119	435	108	488	121	-15	-30	0.32
Midwest	Public	Middle Size City	475	105	500	117	476	103	520	117	-1	-20	1.04
Midwest	Public	Small City	496	100	531	104	497	101	573	99	-1	-42	0.80
Midwest	Catholic	State	475	102	504	103	455	98	481	98	20	23	0.74
Midwest	Public	Small City	522	96	560	105	479	107	533	115	43	27	0.76
Midwest	Catholic	Large City	498	101	524	105	478	104	517	113	20	7	0.54
Midwest	Public	Large City	454	117	497	128	449	110	516	125	5	-19	0.62
Midwest	Public	Middle Size City	479	102	528	118	466	108	535	115	13	-7	0.82
Midwest	Public	Middle Size City	492	102	541	110	472	102	549	110	20	-8	0.63
Midwest	Public	Small City	420	105	446	103	395	103	426	106	25	20	1.00
West/Southwest	Public	Large City	486	105	514	116	465	113	506	117	21	8	0.56
West/Southwest	Public	Small City	451	106	488	104	437	106	465	111	14	23	1.27
West/Southwest	Public	Small City	461	101	519	105	443	99	498	106	18	21	1.05
West/Southwest	Public	Middle Size City	465	115	494	118	457	109	482	115	8	12	0.94
West/Southwest	Public	Large City	471	105	509	108	456	110	507	111	15	2	1.28
West/Southwest	Public	Middle Size City	418	107	445	113	390	109	421	119	28	24	0.62
West/Southwest	Public	Small City	427	110	464	115	399	112	441	116	28	23	0.88

(continued)

Table 2 (continued)

			Class of 1971-72				Class of 1975-76				1972 to 1976 Mean Diff.		
			SAT-V		SAT-M		SAT-V		SAT-M				
Region	Type	Location	M	SD	M	SD	M	SD	M	SD	SAT-V	SAT-M	N(76)/N(72)
West/Southwest	Public	Middle Size City	456	112	480	114	430	106	477	112	26	3	0.83
West/Southwest	Public	Small City	457	103	474	103	444	106	495	104	13	-21	0.91
West/Southwest	Public	Large City	438	110	462	112	412	110	430	114	26	32	0.87
West/Southwest	Public	Small City	495	107	526	104	454	110	507	118	41	19	0.90
West/Southwest	Public	Small City	462	102	507	110	445	91	505	98	17	2	0.91
West/Southwest	Public	Small City	417	105	444	109	416	102	456	115	1	-12	1.19
West/Southwest	Public	Small City	486	102	516	110	459	102	507	113	27	9	1.05
West/Southwest	Public	Small City	462	98	487	106	432	95	469	106	30	18	1.33
West/Southwest	Public	Small City	469	100	501	102	442	97	486	113	27	15	0.75
West/Southwest	Public	Middle Size City	438	106	468	112	430	109	473	113	8	-5	0.75
West/Southwest	Public	Small City	455	104	488	113	449	113	498	117	6	-10	0.63
West/Southwest	Public	Small City	448	96	485	107	452	102	496	112	-4	-11	0.66
West/Southwest	Public	Small City	461	102	501	103	445	98	493	102	16	8	1.08
West/Southwest	Public	Small City	402	104	444	105	378	112	427	123	24	17	0.80
West/Southwest	Public	Middle Size City	470	101	502	105	457	102	503	109	13	-1	1.19
West/Southwest	Public	Small City	336	87	362	86	297	74	344	80	39	18	1.44
West/Southwest	Public	Small City	486	101	517	104	457	101	497	111	29	20	1.26
West/Southwest	Public	Small City	473	104	510	111	442	99	482	111	31	28	1.68
West/Southwest	Public	Small City	467	100	510	109	444	106	495	110	23	15	0.88
West/Southwest	Public	Middle Size City	439	121	466	118	449	122	477	122	-10	-11	0.46
West/Southwest	Public	Small City	465	108	490	111	456	106	495	118	9	-5	0.74
West/Southwest	Public	Middle Size City	498	107	530	112	486	104	528	112	12	2	0.86
West/Southwest	Public	Middle Size City	486	108	512	113	473	106	506	121	13	6	0.49
West/Southwest	Public	Large City	444	109	472	113	413	114	444	116	31	28	0.85
West/Southwest	Public	Large City	414	110	456	116	369	106	416	108	45	40	0.99
West/Southwest	Public	Middle Size City	499	105	515	111	489	105	506	120	10	9	0.78
West/Southwest	Public	Large City	431	114	457	115	410	113	438	112	21	19	0.82
West/Southwest	Public	Middle Size City	415	109	441	108	391	107	427	112	24	14	0.90
West/Southwest	Public	Small City	409	103	436	110	391	105	413	104	18	23	0.96

The Stability of the SAT-Verbal Score Scale

Christopher C. Modu
June Stern
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

May 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $3 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

ACKNOWLEDGMENTS

Grateful acknowledgments are due Bruce K. Eckland, William B. Schrader and Ledyard R Tucker for their invaluable suggestions for improving the design of the study, to Rebecca Bullock for her significant contribution to the completion of the analysis phase of the study, and to William H. Angoff, George H. Hanford, Rex Jackson, Frederick M. Lord, Gary L. Marco, and E. Elizabeth Stewart for their careful reviews of the initial draft of this report.

This study was funded by the Joint Staff Research and Development Committee of the College Entrance Examination Board and Educational Testing Service.

ABSTRACT

This study was designed to determine the extent of drift in the Scholastic Aptitude Test (SAT) score scale between 1963 and 1973 for the SAT-verbal sections only. In a previous study, scores on two old forms were equated to a 1973 form and, through that form, to the College Board scale by means of sets of items common to the new and old forms. In the present study, pairs of 1963 and 1973 forms of the verbal sections were administered in counterbalanced order to spaced samples of the same group, each candidate taking a 1963 and a 1973 form. The obtained scores were used to place the 1973 scores on the reporting scale used for the 1963 form. The experimentally derived scores on the 1963 scale were then compared with their corresponding scores on the 1973 scale for candidates of the same ability levels in order to estimate the degree of scale shift.

The findings of the present study confirm an upward scale drift that gave the 1973 candidate group an average of 8 to 10 points higher than they would have earned had these experimental equating results been used in reporting the 1973 scores rather than the operational equating results. The scale drift observed for the present study was not uniform over the extent of the scale; it was found to increase as scores decreased from 600 to 200. Little scale drift was noticed over the 10-year span for scores of 650 and above. On the basis of the evidence from this study it would appear that operational equating underestimates the mean verbal score declines by about 8 to 10 points over the 10-year span covered by the study. If one of the two conversion lines for the May 1963 operational equating is ignored because of the inconsistency of its results, and if only the second conversion line is considered along with the results for the December 1963 equatings, the scale drift would be about 11 to 13 points rather than 8 to 10.

INTRODUCTION

Considerable interest continues to be expressed in the stability of the verbal and mathematical scales of the Scholastic Aptitude Test. Test-score recipients wish to be reassured that a given scaled score represents the same level of performance on different forms of the SAT. Since scores on each test are placed on scale through form-to-form equating, only by establishing the consistency of the scale system or determining the extent of drift in the scale over a period of time are users able to make meaningful comparisons of populations tested during those years.

The present study was designed to ascertain the extent of drift that may have occurred in the verbal score scale between 1963 and 1973. The study was limited to the verbal sections of the test partly because the mean SAT score decline during the 10-year period covered by the study was greater for verbal than for mathematical scores and partly because the content specifications of the 1976 operational mathematical test differ from those of the 1973 test and it was therefore felt that the candidates used for the study might have been ill-prepared to handle the 1963 or 1973 mathematical test.

In a recent study of the stability of the SAT score scale, Modu and Stern (1975) concluded that the scale had not remained invariant over the 10-year period, 1963-1973. The results of their study showed that the candidates who took the December 1973 form of the SAT earned higher scores than they would have obtained had they taken the December 1963 or 1966 form. The amount of the difference was estimated to be 14 points on the verbal test and 17 points on the mathematical test. Although Modu and Stern followed the same method used by Stewart (1966) in her study--which reached the conclusion that the verbal scale was relatively stable between 1953 and 1963--their report underscored the following methodological considerations that might have contributed to the observed discrepancy between the scores obtained by a candidate on the 1963 or 1966 form and the scores a 1973 candidate of equivalent ability would have earned had he taken those forms:

1. The effect of differences in the performance of the old and new form samples;
2. the differences resulting from the use of Tucker and Levine methods (Angoff, 1971) for operational equatings;
3. the effect of the methods of obtaining test reliability estimates used for the Levine equating procedure;
4. the effect of the choice of equating sections; and
5. the effect of the choice of the old form to which the new form is equated.

Thus, both the Stewart and the Modu and Stern studies are subject to the criticism that they used the same methodology employed in operational procedures that may have led to the reported scale shifts. Hence, the present study was designed to verify the SAT-verbal scale drift during the same 10-year period covered by Modu and Stern in their previous study by employing a different experimental procedure from the one used in earlier studies.

METHOD

Design of the Study

The procedure followed in this study involved the experimental equatings of the raw scores on the 1973 forms to those of the 1963 forms. The resulting raw-to-scaled score conversions were then compared with the scaled scores that were derived when the 1973 forms were first administered. If the score scale had remained stable, a given raw score on the 1973 test, for example, would have converted to the same scaled score on the 1963 or 1973 scale.

Unlike the method used by Stewart (1966) or Modu and Stern (1975), in which a recent form was administered along with the common equating sections from two earlier forms, the present study called for the simultaneous administrations of pairings of 1963 and 1973 verbal tests to four spaced samples of students in each participating school in the following order:

Sample A: May 1963 test (Y1); March 1973 test (X1)
Sample B: March 1973 test (X1); May 1963 test (Y1)
Sample C: December 1963 test (Y2); November 1973 test (X2)
Sample D: November 1973 test (X2); December 1963 test (Y2)

Thus, each student took one 1963 and one 1973 test, with each pair of 1963 and 1973 tests given in counterbalanced order to two different samples. The four test forms were, therefore, paired for printing in four test books as Y1-X1, X1-Y1, Y2-X2 and X2-Y2 and then administered in sequence (spiraled) in each school (i.e., as Y1-X1, X1-Y1, Y2-X2, X2-Y2, and repeat).

The experimental equating methods used in the present study for placing scores from the 1973 tests on the 1963 scale are described below. However, since the design of the study did not require the use of common equating sections, the possible biases listed under items (1) to (4) in the preceding section were eliminated. Also, the methodological weakness of the previous study indicated under item (5) was partially controlled by using two 1963 and two 1973 forms for the equatings instead of only one pair of 1963 and 1973 tests. Ideally, one should have averaged the results of equatings based on pairs of all four 1963 and five 1973 forms, had it been administratively possible to give all nine forms to random samples of the same group. Since the various SAT forms are built to parallel specifications--an assumption that will be verified in the present study--the variations in performance attributable to different pairings of 1963 and 1973 tests should be minimal when the tests are administered to random samples of the same group. Thus, the selection of two 1963 and two 1973 tests is designed (1) to replicate the experimental equating of a 1963 and a 1973 form used in the previous study, and (2) to determine the extent of scale drift based on an average of two independent pairings of test forms administered in the 10-year span.

SAMPLES

Sixty-six high schools in Connecticut, New Jersey, and Pennsylvania that test 100 or more candidates for the SAT each spring participated in the study. Virtually all the schools had mean scores that approximated that of the 1973 national SAT candidates. Motivation for taking the SAT in an experimental administration was partially secured by limiting the tests to college-bound juniors who were planning to take the SAT as juniors in June 1976 or as seniors in the fall of 1976 and might, therefore, benefit from the experience. Additional motivation was also assured by providing rosters of scaled scores obtained on the 1973 test to participating schools for release to students within six weeks of the test administration.

A total of 3,174 students participated in the study. Each of them took a 1963 and a 1973 test during a 2-1/2 hour testing period. Although student participation in the selected schools was voluntary, the schools were encouraged to ensure that the sample of about 40 to 60 students from each school represented the full range of abilities within the SAT-taking population for that school.

Between two and four cases were dropped at random from each of the four samples (A, B, C, and D) in order to reduce each sample to a maximum multiple of five cases as required by the item-analysis computer programs. These reduced samples were then used for placing the scores of 1973 verbal tests on the scale of the 1963 tests.

The raw-score summary statistics for the 1973 and 1963 SAT candidates compared to those for the study samples who took corresponding test forms are presented in Table 1. As can be observed from the table, the samples who took the 1973 tests first obtained higher mean scores than the other samples who took the same tests second. A statistically significant difference ($p < .01$) was observed between the means for the samples that took Test X1 first or second but not between those of the samples that took each of Tests X2, Y1, and Y2 first or second. The combined study samples obtained a higher mean score than the 1973 SAT candidates on Test X1, whereas the combined samples obtained a lower (but nonsignificantly different) mean score on Test X2 than the 1973 SAT candidates. In general, the study samples are of comparable or slightly higher ability than corresponding SAT candidate groups on the 1973 tests. By contrast, all four study samples obtained substantially lower mean scores (averaging 4 to 6 raw-score points) on the 1963 tests than the 1963 SAT candidates. These data emphasize the reality of mean score decline over the 10-year span covered by the study, since the study samples are comparable with the 1973 SAT candidates, as has already been pointed out.

Because there was some concern about the representativeness of the experimental samples compared with that of the operational SAT candidate groups, each sample was subdivided into high- and low-ability groups according to their self-reported class ranks. The equatings based on the total experimental groups for each pair of 1963 and 1973 tests were then repeated separately for high- and low-ability groups in order to assess the difference that any possible variation of the experimental sample from national SAT candidate groups might have had on the equating results.

Tests

As indicated earlier, two 1963 and two 1973 verbal tests were reprinted in four test books for the present study. Each test book contained a 1963 and a 1973 form, with each test divided into two separately timed sections of 30 and 45 minutes, thus resulting in an actual testing time of 150 minutes per candidate. There were no changes in either the timing or the number of items for the SAT-verbal during the 10-year period covered by the study nor were there any variations in format, content, and statistical specifications of the tests that could have seriously affected candidate performance. See Appendix I for the content specifications of the verbal tests.

Because the 1963 and 1973 tests were readministered to high school students in 1976, it was considered necessary to include in the design for the study item analysis procedures that might reveal, from the scatterplots of item-difficulty indices observed on the 1963 tests for (1) 1963 SAT candidates, and (2) the 1976 study samples, whether the 1963 test questions are as appropriate for the 1976 study samples. A similar analysis was conducted for the 1973 tests.

Table 1. Raw Scores Obtained by 1973 and 1963 SAT Candidates Compared with Those of Study Samples Who Took Corresponding Verbal Tests

Test form	SAT candidates	Sample taking test first	Sample taking test second	Combined samples
March 1973 (X1)				
N	120,593	790	790	1,580
Mean.	33.1	35.7[a]	34.1	34.9[b]
S.D.	17.2	17.1	16.1	16.6
November 1973 (X2)				
N	377,846	795	785	1,580
Mean.	36.0	36.0	34.9	35.4
S.D.	17.1	16.0	16.1	16.1[b]
May 1963 (Y1)				
N	250,905	790	790	1,580
Mean.	38.2	34.1	34.5	34.3[b]
S.D.	16.8	16.9	17.6	17.2
December 1963 (Y2)				
N	484,705	785	795	1,580
Mean.	41.3	35.1	34.9	35.0[b]
S.D.	17.1	16.8	16.9	16.8

a. Significantly different at $p < .01$ from corresponding statistics for the sample taking test second.

b. Significantly different at $p < .01$ from corresponding statistics for SAT candidates.

Equating Methods

The linear method described fully as Design II by Angoff (1971, p. 573) and as Case III by Lord (1950) was used for the experimental equating. This method assumes that:

1. The first of a pair of tests administered to a sample of examinees may have a practice or fatigue effect on the results obtained for the test taken second. This assumption requires that the pair of tests be administered in counterbalanced order to random halves of the examinees. Adjustments are made for the estimated practice or fatigue effect as part of the equating procedure.
2. The shape of the distribution of raw scores for one test does not differ markedly from the shape of the distribution of scores for the second test.
3. The pair of tests being equated are parallel forms.

After making adjustments for the estimated practice or fatigue effect, the means and standard deviations for the scores earned on each of the two forms were set equal to each other, and the slope and intercept of the line relating the scores on one form (1973) to the scores on the other (1963) were then derived.

In addition to the above linear procedure, a curvilinear method of equating, the equipercentile method (Angoff, 1971, p. 581), was used to equate scores on the pair of tests taken by each of the four samples, A, B, C, and D. The equatings were then averaged for Samples A and B combined in order to place the March 1973 test on the reporting scale for May 1963, and similarly for Samples C and D combined for the November 1973 test. The equipercentile equating results on the 1963 scale for the average of the two 1973 tests provided an additional check on the linear equating results.

Score Equating for High- and Low-Ability Groups

As an added check on our equating results, the linear equating method was applied to subsamples of differing ability levels. By means of a self-reported class rank indicated by the student at the time of the experimental administration (see Appendix VI), the samples were divided into two groups: (1) students who indicated that they stood in the top fifth of their high school class, and (2) students who responded that they were in one of the four categories below the top fifth. Results from these equatings were compared with those obtained previously for all students in each of the four samples.

RESULTS AND DISCUSSION

Appropriateness of the 1963 and 1973 Verbal Items for 1976 Study Samples

The equating results of this study could be affected by differential changes in the difficulty level of some of the items for the 1963 or

1973 SAT candidates versus the 1976 experimental samples. If, for example, some of the items on the 1963 test had become obsolete for the 1976 experimental samples and had changed radically in difficulty from the results obtained when the test was originally taken by the 1963 candidate group, an apparent shift in the score scale could have resulted from the differential changes in item difficulties that could have altered the mean scores for the experimental equating samples. Analyses of the item-difficulty indices were, therefore, carried out to determine whether any items in the 1963 and 1973 tests were inappropriate for the 1976 study samples.

As can be noticed from the scatterplots of the item-difficulty indices (deltas) for the 370-case samples of the 1963 candidate groups versus the 1976 experimental samples (see Appendixes II and III), the delta plots form elongated ellipses with plot correlations of .964 for both the May 1963 and the December 1963 test forms. Plots for the 1070- and the 1995-case samples of the March and November 1973 test forms, respectively, versus the 1976 experimental samples (see Appendixes IV and V) form tighter ellipses than those of Appendixes II and III. Their plot correlations of .989 and .994, which are higher than those of the 1963 tests, may be due to the fact that the 1976 samples are more like the 1973 than the 1963 candidate group.

Items that fall on the major axis of the ellipse are of relatively equal difficulty for both groups; points above the line represent items that are relatively more difficult for the 1963 sample, whereas those below the line represent those which are relatively more difficult for the 1976 study samples. No item type appears to be relatively easier or more difficult for any one group, since all item types straddle both sides of the major axis of the ellipse. One can observe from the displacement of the ellipse of points to the right of the equal-difficulty line for both groups (i.e., the broken line that forms a 45-degree angle with the vertical axis) that the 1976 study samples are not as able as the 1963 candidate groups. The relative difficulty of the 1963 tests for the SAT candidate groups and the study samples is further confirmed by the observed difficulty indices, delta, of 12.6 and 12.4 for the May 1963 and December 1963 groups, respectively, compared to 13.1 and 13.0 for the two study samples, A and C, who took these tests first. In comparison with the 1973 candidate groups, the 1976 study samples are seen from the scatterplot in Appendix IV to be slightly more able than the March 1973 candidate group, whereas the study samples are of comparable ability with the November 1973 candidate group in view of the virtual collinearity of the major axis of the delta plots with the 45-degree, equal-difficulty line.

A couple of items in the plots for 1963 tests that might be considered more difficult for one sample relative to the rest of the items in the test were scrutinized for obsolescence or changes in word meaning over a 13-year span. In each case, no apparent justification was found for the observed differential changes in difficulty. Since the plot correlations are no lower than .96 and relatively few items were found to evidence appreciable differential changes in difficulties, it seems likely that such differential changes have little effect, if any, on either the mean scores of the experimental samples or the equating results.

Equating Results

Raw scores on two pairs of 1963 and 1973 tests that were administered in counterbalanced order were equated by setting means and standard deviations equal after making the necessary adjustments for practice or fatigue effect. As will be noted from the summary statistics in Table 2, the conditions for the linear equating method of this study are satisfied: the four pairs of tests are parallel forms with minimal observed differences in their reliabilities, standard errors of measurement, difficulty indices and distributional characteristics. The raw-score distributions on the pair of tests taken by each sample are given in Appendixes VII to X.

In adjusting for practice or fatigue effect for the Design II equating method, a slight fatigue effect was observed for all four samples. This was estimated to result in a depressed raw-score of .64 to .68 of a point (or about 4 scaled-score points) on the test taken second for each of the samples.

The slopes and intercepts of the operational conversion lines used for placing the March and November 1973 SAT-verbal raw scores on the reporting scale as well as those of the experimentally derived conversion lines for transforming the raw scores on these tests to their equivalent scores on the May and December 1963 scales, respectively, are presented in Table 3. For a given reported score on the March or November 1973 test, the equivalent score on the corresponding May or December 1963 scale was computed from these conversion parameters. The two experimentally derived scores on the 1963 scales were then averaged to yield a one-to-one correspondence between a 1973 reported score and its equivalent score on the 1963 scale.

Presented in Table 4 are the scores at specific points on the 1973 reported-score scale and their equivalent linear and curvilinear experimentally derived scores on a 1963 scale for each of the 1973 tests, as well as for the average of the two 1973 tests. The linear equating results indicate that a performance level of 800 as reported in 1973 corresponds to an average of 803 on the rederived scale for the two 1973 forms; similarly, a reported score of 600 corresponds to 596 on the experimentally derived scale; a reported score of 450 corresponds to 441; and a reported score of 300 corresponds to 286.

Thus, at the reported score levels of 600 to 800 there was a maximum possible shift in scale of 4 points. However, as the reported score level decreases below 600, the scale shift increases to a maximum of 18 points for the average of the two forms at the 200 score level. These results are illustrated in the graph of the 1973 reported scores compared with that of their equivalent scores on the experimentally derived 1963 scale (see Appendix XI).

As presented in Table 5, the reported mean score of 453 for the four experimental samples corresponds to a derived mean score of 444 for a scale shift of 9 points over the 10-year period for the average of the two experimental equatings. That is, the mean reported score was 9 points higher, as determined from the study, than it would have been had the scale remained invariant. The actual candidate groups tested in March and November 1973, totaling 500,965 candidates, obtained an average reported mean score of 454 compared to 453 for

Table 2. Summary Statistics for Four Samples (A-D)
That Took Pairings of 1963 and 1973 SAT-Verbal in Counterbalanced Order[a]

			Raw scores						Standard error of measurement	Difficulty index (delta)[c]		R-Biserial	
Sample	N	SAT-verbal test[a]	Mean	S.D.	Range	Skewness	r_{xy}[b]	Reliability (K-R#20)		Mean	S.D.	Mean	S.D.
A . . .	790	May 1963 (Y1)	34.1	16.9	-10-86	0.15	.90	.92	4.9	13.1	2.6	0.46	0.11
		March 1973 (X1)	34.1	16.1	- 5-80	0.31	(.98)	.91	4.8	13.1	2.7	0.45	0.12
B . . .	790	March 1973 (X1)	35.7	17.1	- 6-83	0.28	.92	.92	4.8	12.9	2.7	0.47	0.12
		May 1963 (Y1)	34.5	17.6	- 2-83	0.29	(1.00)	.92	4.9	13.0	2.6	0.48	0.11
C . . .	785	December 1963 (Y2)	35.1	16.8	- 5-83	0.17	.90	.92	4.9	13.0	2.6	0.46	0.12
		November 1973 (X2)	34.9	16.1	- 5-80	0.24	(.98)	.91	4.8	13.0	2.8	0.45	0.13
D . . .	795	November 1973 (X2)	36.0	16.0	- 1-79	0.22	.90	.91	4.8	12.9	2.8	0.44	0.13
		December 1963 (Y2)	34.9	16.9	- 4-81	0.23	(.98)	.92	4.9	13.0	2.5	0.45	0.11

a. Presented in the order in which the tests were actually taken.

b. Correlation between scores on the 1963 and 1973 tests, with the correlation corrected for attenuation in parentheses.

c. The difficulty index, delta, is the percent correct for each item expressed as a normal deviate on a linear scale with a mean of 13.0 and a standard deviation of 4.0. The higher the delta value, the more difficult the item.

Table 3. Parameters for Placing 1973 SAT-Verbal Scores on the 1963 and 1973 Scales

	Conversion parameters				
	1973 scale (operational)		1963 scale (experimental)		Possible scaled-score range
Test form	Slope	Intercept	Slope	Intercept	
March 1973	6.3059	231.6546			93-799
			6.5404	218.7723	75-807
November 1973 . . .	6.3281	229.9625			91-799
			6.5219	209.5446	66-797

the experimental samples. One can, therefore, conclude that the experimental samples represent the same ability level as the actual 1973 candidate groups.

If the two pairs of equating results had not been averaged but had been considered separately, it can easily be observed from Table 4 that the results for the November 1973 test produced a larger scale shift than was evidenced from the March 1973 equating results. Thus, a discrepancy in the equating lines resulting from the choice of a specific pair of forms to be equated does exist. As indicated above in the design of the study, the ideal method of establishing the scale shift between 1963 and 1973 would have been to equate the scores on all possible pairings of the four 1963 and five 1973 forms and to have averaged the results for the 20 pairings of equatings that would have been administered to 40 samples in counterbalanced order.

Had the May 1963 and the March 1973 forms only been used for the experimental equatings, it would have been concluded (see Table 4) that an average score of about 453 on the 1973 scale had shifted upward from the 1963 scale by about 5 points. By comparison, the experimental equating of the December 1963 and November 1973 test forms resulted in an upward scale shift of about 14 points at the same 1973 reported-score level of 453.

The results of the curvilinear equatings, also presented in Table 4, closely parallel the linear results. Comparisons of the average of the linear with the average of the curvilinear equating results for the two 1973 forms reveal a discrepancy of at most 5 points at the extremities of the score scales (ignoring the extrapolated values in parentheses) but no difference at the 500 score level. The linear and the equipercentile results are consistent in the middle range of the score scale.

Table 4. Comparison of 1973 Reported Scores with Equivalent Scores on the 1963 Experimental Derived Scale

1973 reported score	Equivalent score on 1963 scale for Form X1 (March 1973 test)		Equivalent score on 1963 scale for Form X2 (November 1973 test)		Equivalent score on 1963 scale for average of two 1973 tests	
	Linear scale	Curvilinear scale	Linear scale	Curvilinear scale	Linear scale	Curvilinear scale
800	808	(785)[a]	797	(781)[a]	803	(783)[a]
750	756	(734)[a]	746	(732)[a]	751	(733)[a]
700	705	700	694	688	699	694
650	653	654	642	639	648	646
600	601	599	591	589	596	594
550	549	550	539	545	544	547
500	497	496	488	488	492	492
450	445	448	436	438	441	443
400	393	397	385	383	389	390
350	342	339	333	332	337	335
300	290	288	282	281	286	285
250	238	233	230	225	234	229
200	186	(194)[a]	179	(182)[a]	182	(188)[a]

a. Figures in parentheses were obtained by extrapolation from graphed data.

Table 5. Summary Statistics on 1973 Tests for Reported vs Experimentally Derived Score Scales for the Study Samples

Test form	Sample[a]	Size of sample	1973 reported score Mean	1973 reported score Std. Dev.	Experimentally derived linear scale Mean	Experimentally derived linear scale Std. Dev.
March 1973	A	790	446.5	101.6	441.6	105.4
	B	790	456.9	107.8	452.4	111.8
November 1973 . .	C	785	450.6	101.7	436.9	104.8
	D	795	457.8	101.4	444.4	104.5
Average of two tests	A-D	3,160	453.0	103.1	443.8	107.0

a. Samples A and C took the 1963 test first and the 1973 test second; samples B and D took the 1973 test first and the 1963 test second.

Discrepancy Between the 1963 and the March and November 1973 Scales

It seems appropriate at this point to explore in greater detail the discrepancy between the average scale drifts of about 5 and 14 points reported for the March 1973 and November 1973 tests, respectively, by examining the results of the operational equating for the May and December 1963 tests to which the 1973 tests used in this study were experimentally equated. In form-to-form equating of operational tests, each new form of the SAT is equated to two old forms. The slopes and intercepts of the conversion lines produced from the two independent equatings are then averaged to yield one conversion line for linking the new form to the reporting scale for the old forms. Usually, the conversion lines from the two independent equatings should be quite close to each other in the middle of the score scale, thus providing an additional check on the consistency of the equating results. Comparisons of the May and December 1963 equating results are presented in Table 6.

At the reported-score level of 453 for the November 1973 test, the two conversion lines with which the December 1963 test was placed on the reporting scale produced equivalent scores of 440 and 439 on the scales for a 1960 and a 1962 form, respectively. However, at the same reported-score level of 453 for the March 1973 test, the two equating lines linking the May 1963 test to two earlier forms yielded equivalent scores of 454 and 443, respectively, on the scales for a 1960 and a 1962 form for a discrepancy of 11 points. The greater consistency between the results of the two independent equating lines for the December 1963 scale and the closeness of those results with the May 1963 (Line 2) equating rather than with the May 1963 (Line 1) equating suggest that the extent of the scale drift ought to be based

Table 6. Comparison of 1973 Reported Scores with Equivalent Scores on the May 1963 and the December 1963 Experimentally Derived Scales Based on Two Separate Equating Lines for Each 1963 Form

1973 reported score	Equivalent score on May 1963 scale for Form X1 (March 1973 test)			Equivalent score on December 1963 scale for Form X2 (November 1973 test)			
	May 1963 equating to 1960 form (Line 1)	May 1963 equating to 1962 form (Line 2)	May 1963 scale from bisector of Lines 1&2	Dec. 1963 equating to 1960 form (Line 1)	Dec. 1963 equating to 1962 form (Line 2)	Dec. 1963 scale from bisector of Lines 1&2	Equivalent score on 1963 scale[a] for average of two 1973 tests
800	824	793	808	785	810	797	803
750	771	743	756	735	757	746	751
700	717	692	705	686	703	694	699
650	664	642	653	636	649	642	648
600	611	592	601	586	596	591	596
550	557	541	549	537	542	539	544
500	504	491	497	487	489	488	492
450	451	440	445	437	435	436	441
400	397	390	393	388	382	385	389
350	344	339	342	338	328	333	337
300	291	289	290	288	275	282	286
250	237	238	238	239	221	230	234
200	184	188	186	189	168	179	182
Mean 453	454	443	448	440	439	439	444

a. Average of the two bisector lines from equivalent scores on May and December 1963 scales.

on the May 1963 (Line 2) and the December 1963 (Lines 1 and 2) equatings, which produced more consistent results. Consequently, the larger upward scale drift of about 14 points produced for the November 1973 scale with respect to the December 1963 equatings appears to be a more reliable estimate than the 5-point shift observed for the March 1973 scale at the reported mean-score level of 453 for the 1973 candidate group.

If the equating results for conversion Line 1 of the May 1963 test were ignored and the reporting of the May 1963 score based only on the second conversion line, the reported score of 453 on the 1973 scale would have been equivalent to an average score of 441 on the 1963 scale. Thus, the upward drift for the 1973 scale would have amounted to an average of 12 points instead of the 9-point difference between 453 and its equivalent score, on the 1963 scale, of 444 produced by using both pairs of 1963 operational equating lines.

Equating Results for High- and Low-Ability Groups

In order to examine whether the equating results based on the experimental samples would vary if their average ability levels had been different from those of actual SAT candidate groups, additional equatings were performed for two subsamples of each experimental group. Each of the four experimental samples was categorized into two subsamples on the basis of their self-reported class ranks as indicated in Table 7. It is observed that the mean score difference between the high- and low-ability groups is approximately one standard deviation of the scores for the total group for each test. Despite the substantial difference between the two ability levels, the conversion parameters derived from the separate equatings based on the performance of high- and low-ability groups produced results similar to those obtained for the total group.

The raw-score means and standard deviations for the total group in each of the four samples were then converted to the experimentally derived scales by applying the parameters obtained from the three separate equatings. The resulting converted scores and parameters are presented in Table 8. It is noted that within each sample, the total-group mean scores and standard deviations based on the different parameters do not vary by more than 2 points. It is, therefore, reasonable to conclude that our study results would not have been affected had the ability levels of the SAT candidate groups been higher or lower than those of the experimental groups (who, incidentally, were found not to have been markedly different from the combined group of actual SAT candidates who took the March and November 1973 SAT-verbal).

Equating Error

In interpreting the observed scale shift between the 1973 reported and the 1963 experimentally derived score scales, consideration must be given to the amount of sampling error in the equating operation. The sampling errors for equating (Lord, 1950) associated with selected scores on the 1973 reporting scale are presented in Table 9 for

Table 7. Summary Raw-Score Data on 1963 and 1973 Verbal Tests for Total vs. High- and Low-Ability Groups[a]

Sample[b]	Ability level	N	Mean	Std. Dev.	Mean	Std. Dev.
			May 1963 test		March 1973 test	
A	High	319	44.8245	14.6210	44.4138	14.0062
	Low	456	27.3355	13.9386	27.4671	13.3176
	Total group	790	34.1241	16.9279	34.0684	16.1157
B	High	335	44.7075	15.7873	45.7701	15.3877
	Low	444	26.9167	14.9392	28.3874	14.1779
	Total group	790	34.4848	17.6199	35.7266	17.0928
			December 1963 test		November 1973 test	
C	High	315	45.3778	14.4585	44.7524	13.9944
	Low	457	28.3895	14.3585	28.4573	13.4222
	Total group	785	35.0803	16.8038	34.8599	16.0744
D	High	319	45.7022	14.7611	45.8777	14.0219
	Low	467	27.7152	14.0154	29.4261	13.6023
	Total group	795	34.9358	16.8781	36.0101	16.0283

a. High are those students who reported their class rank to be in the top fifth. Low are those students who reported their class rank to be below the top fifth. Total group includes those students who did not indicate their class rank.

b. Samples A and C took the 1963 test first; samples B and D took the 1973 test first.

the linear equating of the March and November 1973 forms to the May and December 1963 forms, respectively. The sampling errors associated with the equivalent scores on the 1963 linear scale for the average of the two 1973 tests (see Table 4) are also presented in the last column of Table 9.

On the basis of these estimated errors, a reported score of 750 on the 1973 scale is equivalent to a scaled score ranging from about 748 to 753 (i.e., 751 ± 2.5) on the 1963 scale for a discrepancy of -2 to +3 points at that level between the 1973 and 1963 scales. This observed discrepancy represents hardly any shift at that score level. Similarly, a reported score of 600 on the 1973 scale is equivalent to a scaled score ranging from about 594 to 597 on the 1963 scale for a discrepancy of 3 to 6 points in favor of the candidate of this verbal ability level who took the 1973 test. Applying the estimated error of about one score point at the 500 score level on the 1973 reporting scale yields an equivalent score of 492 ± 0.9 on the 1963 scale for a scale shift of 7 to 9 points in favor of a 1973 candidate of the same ability. Similarly, a mean score of 453 on the 1973 reporting scale, which is equivalent to an experimentally derived score of 444 on the 1963 scale, yields a discrepancy of 8 to 10 points between the two scales when the equating error of about 1 score point at this level is taken into account. At the 300 score level, the discrepancy or shift between the 1973 and 1963 scales ranges from about 13 to 16 points in favor of the 1973 candidate. Thus, the results in Table 4 indicate almost no scale drift at the 650 score level and above, but drifts ranging from 2 scaled-score points at the 600 verbal score ability level to almost 20 score points at the 200 score level in favor of the 1973 candidates compared to the scores that 1963 candidates of corresponding ability levels had earned on the 1963 SAT-verbal.

Table 8. Conversion Parameters Derived from Separate Equating Results for Different Ability Subgroups and the Corresponding Total-Group Statistics Based on Those Parameters

Sample	Group used for parameters	Experimentally derived conversion line[a]		Total-group statistics using different parameters		Possible scaled-score range
		Slope	Intercept	Mean	Std. Dev.	
A . . .	High ability	6.5029	219.5720	441	105	77-805
	Low ability	6.6052	217.4121	442	106	72-812
	Total group	6.5404	218.7723	442	105	75-807
B . . .	High ability	6.5029	219.5720	452	111	77-805
	Low ability	6.6052	217.4121	453	113	72-812
	Total group	6.5404	218.7723	452	112	75-807
C . . .	High ability	6.4834	212.3194	438	104	70-796
	Low ability	6.5268	208.5140	436	105	65-796
	Total group	6.5219	209.5446	437	105	66-797
D . . .	High ability	6.4834	212.3194	446	104	70-796
	Low ability	6.5268	208.5140	444	105	65-796
	Total group	6.5219	209.5446	444	105	66-797

a. Conversion parameters for corresponding ability groups in Samples A and B, and C and D are identical because Design II linear equating combines group statistics for counterbalanced pairings of tests.

Table 9. Standard Error[a] of Equated Score

1973 SAT-verbal reported score	Experimentally derived verbal score on 1963 scale		
	March 1973 test	November 1973 test	Average for March and November 1973 tests
750	756 ± 3.5	746 ± 3.6	751 ± 2.5
600	601 ± 2.0	591 ± 2.1	596 ± 1.5
500	497 ± 1.3	488 ± 1.3	492 ± 0.9
400	393 ± 1.3	385 ± 1.3	389 ± 0.9
300	290 ± 2.0	282 ± 2.1	286 ± 1.5
200	186 ± 3.0	179 ± 3.2	182 ± 2.2

a. The estimated error due to sampling variance.

CONCLUSION

The results of this study support the conclusion reached in the 1975 study by the present authors that the SAT-verbal scale has drifted upward over the 10-year period, 1963-1973, but the magnitude of the drift is less than was estimated in the previous study. The experimental samples obtained a mean score of 453 on the 1973 verbal test forms that was found to be about 9 points higher than the mean score of 444 they would have earned on the 1963 reporting scale. It was observed, however, that little or no scale shift occurred in the score range of 650 and above.

The 9-point scale shift was also found to be smaller than the 16-point upward shift obtained in the previous study at the 1973 reported-score level of 453, which was then determined (from the results of the common-item equating of the December 1963 test form to that of December 1973) to be equivalent to 437 on the 1963 scale. Had the two independent equatings of the present study not been averaged, the scale shift at the 1973 mean-score level of 453 would have amounted to an upward drift of only about 5 points for the equating of one pair and about 14 points for the equating of the other pair of tests. Thus, there is an obvious difference in the equating results associated with the particular pair of tests selected for form-to-form equating. If one of the two conversion lines from the May 1963 operational equating that appears to provide inconsistent equating results with the second conversion line were ignored and only the line that provides more consistent results with those of the December 1963 equatings were considered, the scale drift would be about 11 to 13 points (taking into account a sampling error of about 1 point) rather than 8 to 10.

The implication of this study for the SAT-verbal score decline is that operational equatings during the 1963-1973 period have had the cumulative effect of making the 1973 candidate groups appear more able than the 1963 candidates of comparable ability. Thus, the study results suggest that the 1973 reported scores underestimate the extent of mean verbal score decline over the 10-year period by about 8 to 10 points, although there is evidence that the scale shift could be as high as 11 to 13 points.

REFERENCES

Angoff, W. H., "Scales, Norms and Equivalent Scores," in R. L. Thorndike (ed.) Educational Measurement (Second Edition). Washington, D.C.: American Council on Education, 1971

Lord, F. M., Notes on Comparable Scales for Test Scores. Research Bulletin 50-48. Princeton, N.J.: Educational Testing Service, 1950

Modu, C. C., and Stern, J., The Stability of the SAT Score Scale. College Entrance Examination Board Research and Development Report 74-75, No. 3. Princeton, N.J.: Educational Testing Service, 1975

Stewart, E. E., The Stability of the SAT-Verbal Score Scale. College Entrance Examination Board Research and Development Report 66-7, No. 3. Princeton, N.J.: Educational Testing Service, 1966.

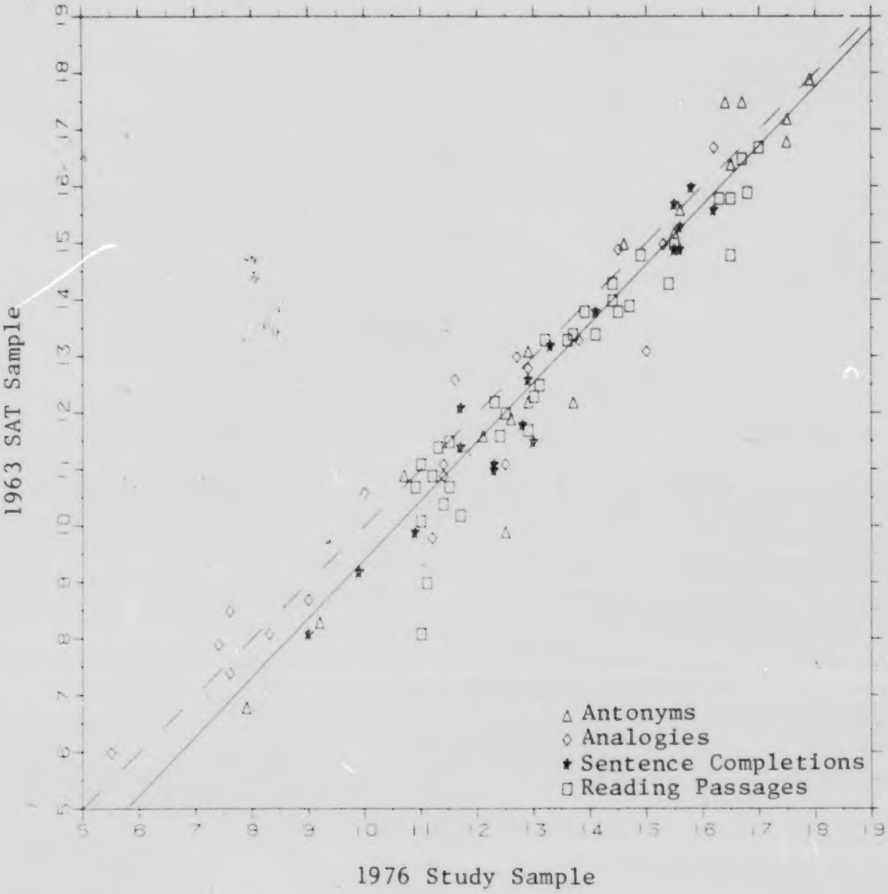
Appendix I. Description of 1963 and 1973 SAT-Verbal Sections on Which Scores Were Reported

Section	Item types	Number of items	
30-minute	Antonyms	10	40
	Analogies	10	
	Sentence Completion	10	
	Reading Passages	10	
45-minute	Antonyms	8	50
	Analogies	9	
	Sentence Completion	8	
	Reading Passages	25	
Total test.			90

Note: All test forms were formula-scored (R-W/4).

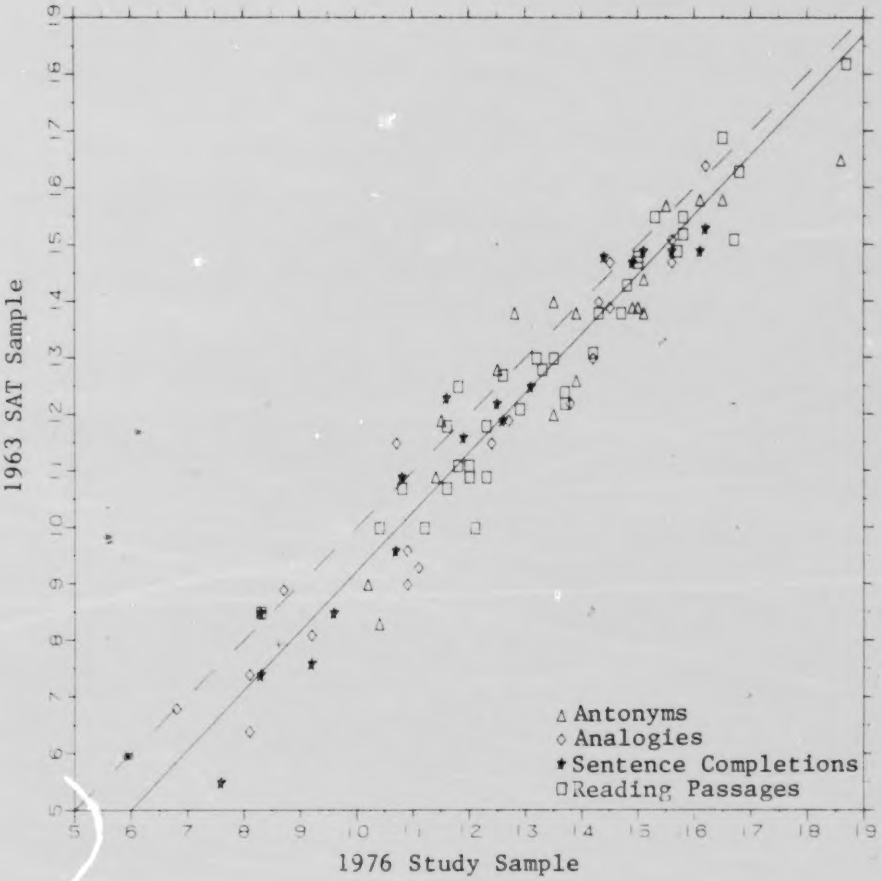
Appendix II. Delta Plot of May 1963 Verbal Items
for 1963 SAT Sample vs 1976 Study Sample

Sample A
R = 0.964



Appendix III. Delta Plot of December 1963 Verbal Items
for 1963 SAT Sample vs 1976 Study Sample

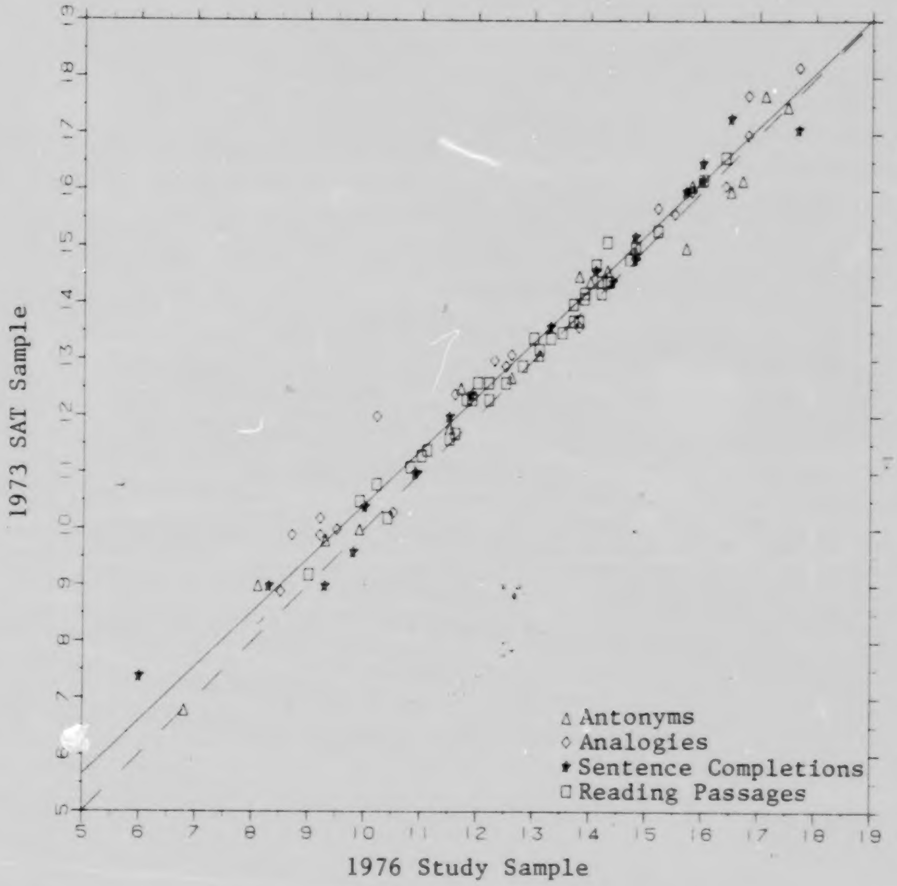
Sample C
R = 0.964



678

Appendix IV. Delta Plot of March 1973 Verbal Items for 1973 SAT Sample vs 1976 Study Sample

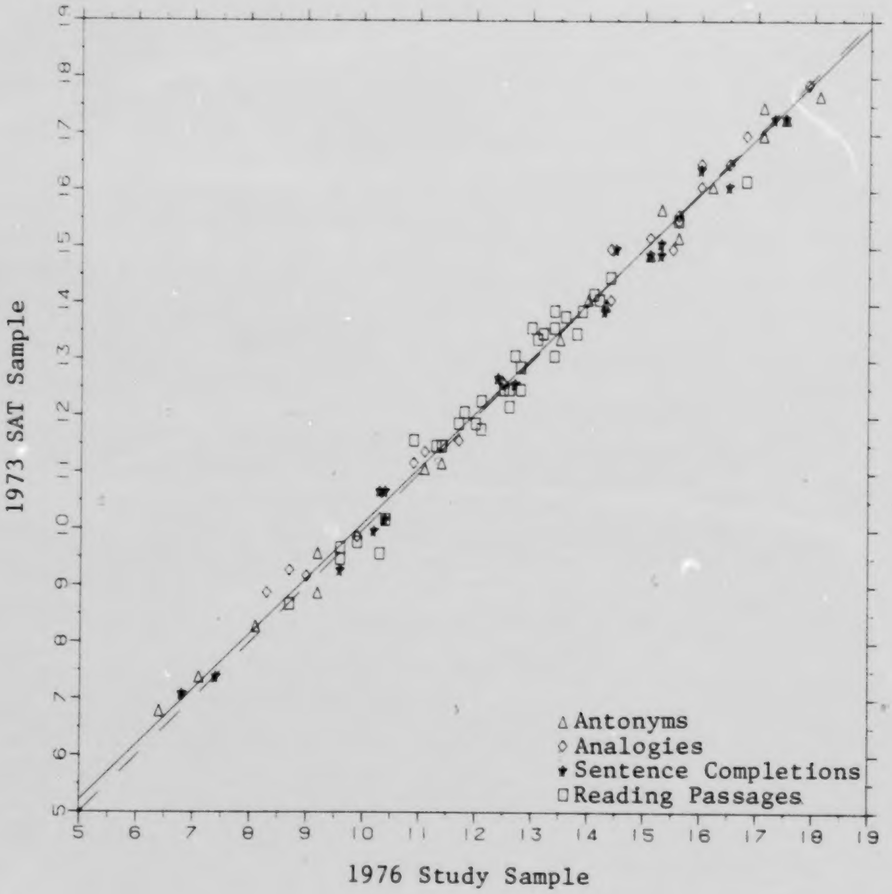
Sample B
R = 0.989



679

Appendix V. Delta Plot of November 1973 Verbal Items for 1973 SAT Sample vs 1976 Study Sample

Sample D
R = 0.994



Appendix VI. Question on Class Rank and the Frequencies of Students in Each Response Category

Question: How would you compare your achievement in subjects taken since beginning ninth grade with that of other students in your high school class?
(A) Top fifth
(B) Second fifth
(C) Third fifth
(D) Fourth fifth
(E) Lowest fifth

Frequencies of Obtained Responses

	Study samples			
Class rank	A	B	C	D
Top fifth	319	335	315	319
Second fifth	256	242	250	250
Middle fifth	167	170	175	175
Fourth fifth	27	28	26	36
Lowest fifth	6	4	6	6
No response	17	15	17	13
Total number	792	794	789	799

Appendix VII. Raw-Score Frequency Distributions of 1963 and 1973 Verbal Tests[a]

Sample A

May 1963 form

Raw score X	f	Percentile rank of lower limit of interval
85. . . . 86	1	99.9
80. . . . 84	0	99.9
75. . . . 79	4	99.4
70. . . . 74	10	98.1
65. . . . 69	17	95.9
60. . . . 64	29	92.3
55. . . . 59	36	87.7
50. . . . 54	60	80.1
45. . . . 49	54	73.3
40. . . . 44	85	62.5
35. . . . 39	83	52.0
30. . . . 34	83	41.5
25. . . . 29	91	30.0
20. . . . 24	72	20.9
15. . . . 19	54	14.1
10. . . . 14	56	7.0
5. . . . 9	32	2.9
0. . . . 4	16	0.9
- 5. . . .- 1	5	0.3
-10. . . .- 6	2	0.0
	790	

March 1973 form

Raw score X	f	Percentile rank of lower limit of interval
80. . . .	2	99.7
75. . . . 79	3	99.4
70. . . . 74	7	98.5
65. . . . 69	22	95.7
60. . . . 64	24	92.7
55. . . . 59	37	88.0
50. . . . 54	54	81.1
45. . . . 49	61	73.4
40. . . . 44	63	65.4
35. . . . 39	78	55.6
30. . . . 34	101	42.8
25. . . . 29	106	29.4
20. . . . 24	85	18.6
15. . . . 19	61	10.9
10. . . . 14	46	5.1
5. . . . 9	26	1.8
0. . . . 4	11	0.4
- 5. . . .- 1	3	0.0
	790	

May 1963 form:

$M_x = 34.12$

$\sigma_x = 16.93$

$Md_x = 33.61$

$n = 90$ items

March 1973 form:

$M_x = 34.07$

$\sigma_x = 16.12$

$Md_x = 32.50$

$n = 90$ items

a. Ordered as taken by the sample.

Appendix VIII. Raw-Score Frequency Distributions of 1973 and 1963 Verbal Tests[a]

Sample B

March 1973 form			May 1963 form		
Raw score X	f	Percentile rank of lower limit of interval	Raw score X	f	Percentile rank of lower limit of interval
80. . . . 83	7	99.1	80. . . . 83	2	99.7
75. . . . 79	4	98.6	75. . . . 79	11	98.4
70. . . . 74	11	97.2	70. . . . 74	12	96.8
65. . . . 69	26	93.9	65. . . . 69	24	93.8
60. . . . 64	31	90.0	60. . . . 64	28	90.3
55. . . . 59	35	85.6	55. . . . 59	39	85.3
50. . . . 54	54	78.7	50. . . . 54	45	79.6
45. . . . 49	62	70.9	45. . . . 49	63	71.6
40. . . . 44	77	61.1	40. . . . 44	67	63.2
35. . . . 39	99	48.6	35. . . . 39	79	53.2
30. . . . 34	80	38.5	30. . . . 34	94	41.3
25. . . . 29	80	28.4	25. . . . 29	94	29.4
20. . . . 24	89	17.1	20. . . . 24	67	20.9
15. . . . 19	56	10.0	15. . . . 19	57	13.7
10. . . . 14	34	5.7	10. . . . 14	51	7.2
5. . . . 9	27	2.3	5. . . . 9	36	2.7
0. . . . 4	15	0.4	0. . . . 4	17	0.5
- 5. . . .- 1	2	0.1	- 2. . . .- 1	4	0.0
- 6. . . .	1	0.0			
	790			790	

March 1973 form	May 1963 form
$M_x = 35.73$	$M_x = 34.48$
$\sigma_x = 17.09$	$\sigma_x = 17.62$
$Md_x = 34.98$	$Md_x = 32.95$
n = 90 items	n = 90 items

a. Ordered as taken by the sample.

Appendix IX. Raw-Score Frequency Distributions of 1963 and 1973 Verbal Tests[a]

Sample C

December 1963 form			November 1973 form		
Raw score X	f	Percentile rank of lower limit of interval	Raw score X	f	Percentile rank of lower limit of interval
80. . . . 83	2	99.7	80. . . .	3	99.6
75. . . . 79	4	99.2	75. . . . 79	5	99.0
70. . . . 74	13	97.6	70. . . . 74	8	98.0
65. . . . 69	13	95.9	65. . . . 69	13	96.3
60. . . . 64	31	92.0	60. . . . 64	27	92.9
55. . . . 59	42	86.6	55. . . . 59	43	87.4
50. . . . 54	60	79.0	50. . . . 54	49	81.1
45. . . . 49	74	69.6	45. . . . 49	69	72.4
40. . . . 44	69	60.8	40. . . . 44	80	62.2
35. . . . 39	87	49.7	35. . . . 39	88	51.0
30. . . . 34	79	39.6	30. . . . 34	89	39.6
25. . . . 29	76	29.9	25. . . . 29	88	28.4
20. . . . 24	82	19.5	20. . . . 24	79	18.3
15. . . . 19	63	11.5	15. . . . 19	69	9.6
10. . . . 14	43	6.0	10. . . . 14	34	5.2
5. . . . 9	31	2.0	5. . . . 9	33	1.0
0. . . . 4	13	0.4	0. . . . 4	3	0.6
- 5. . . .- 1	3	0.0	- 5. . . .- 1	5	0.0
	785			785	

$M_X = 35.08$	$M_X = 34.86$
$\sigma_X = 16.80$	$\sigma_X = 16.07$
$Md_X = 34.66$	$Md_X = 34.00$
$n = 90$ items	$n = 90$ items

a. Ordered as taken by the sample.

Appendix X. Raw-Score Frequency Distributions of 1973 and 1963 Verbal Tests[a]

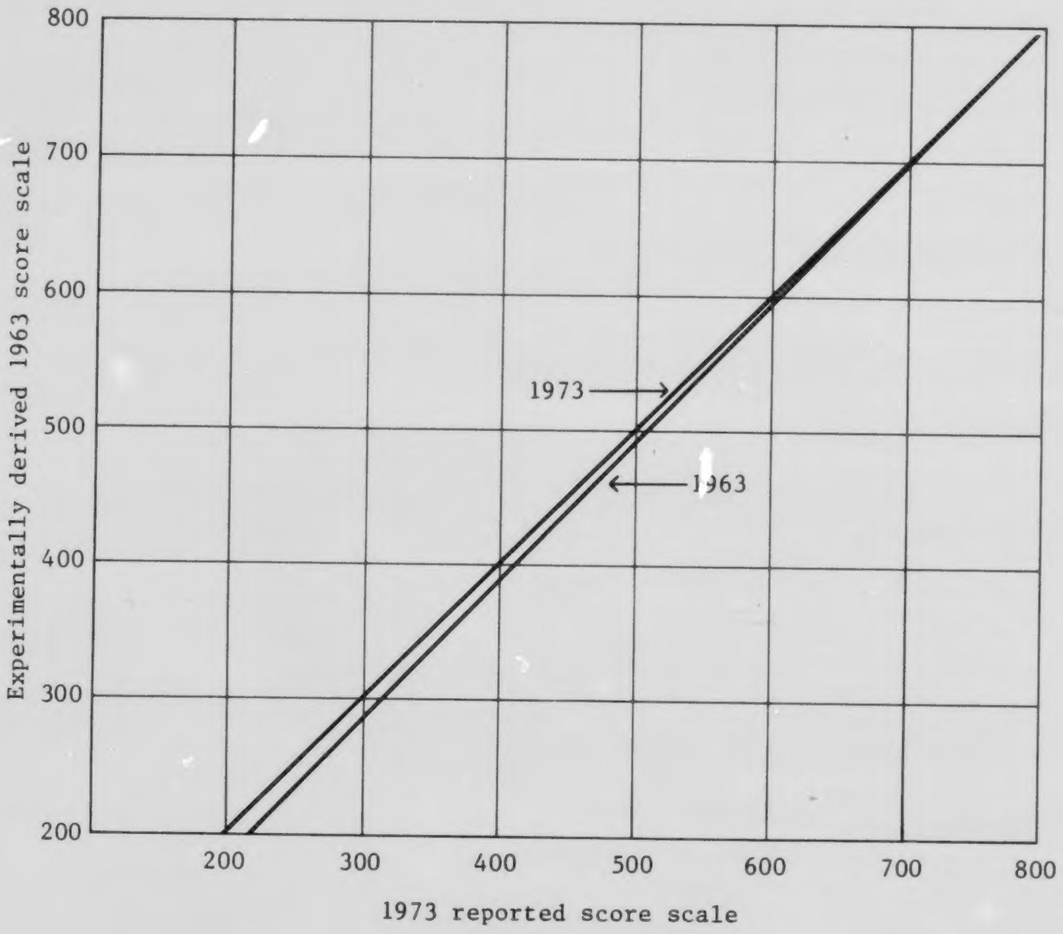
Sample D

November 1973 form			December 1963 form		
Raw score X	f	Percentile rank of lower limit of interval	Raw score X	f	Percentile rank of lower limit of interval
75. . . . 79	6	99.2	80. . . . 81	1	99.9
70. . . . 74	12	97.7	75. . . . 79	10	98.6
65. . . . 69	18	95.5	70. . . . 74	7	97.7
60. . . . 64	28	91.9	65. . . . 69	20	95.2
55. . . . 59	46	86.2	60. . . . 64	36	90.7
50. . . . 54	56	79.1	55. . . . 59	32	86.7
45. . . . 49	78	69.3	50. . . . 54	64	78.6
40. . . . 44	82	59.0	45. . . . 49	67	70.2
35. . . . 39	77	49.3	40. . . . 44	69	61.5
30. . . . 34	88	38.2	35. . . . 39	78	51.7
25. . . . 29	93	26.5	30. . . . 34	90	40.4
20. . . . 24	82	16.2	25. . . . 29	87	29.4
15. . . . 19	64	8.2	20. . . . 24	75	20.0
10. . . . 14	34	3.9	15. . . . 19	68	11.4
5. . . . 9	25	0.8	10. . . . 14	49	5.3
0. . . . 4	5	0.1	5. . . . 9	23	2.4
- 1. . . .	1	0.0	0. . . . 4	17	0.3
			- 4. . . .- 1	2	0.0
	795			795	

$M_x = 36.01$	$M_x = 34.94$
$\sigma_x = 16.03$	$\sigma_x = 16.88$
$Md_x = 34.87$	$Md_x = 33.86$
n = 90 items	n = 90 items

[a] Ordered as taken by the sample.

Appendix XI. 1973 Reported Scores Compared to Their Equivalent Scores on the 1963 Scale



A Review of Data Available regarding Family Income and Financial Aid Characteristics of Students

James E. Nelson
College Board

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

October 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

Introduction

One of the explanations offered for the Scholastic Aptitude Test (SAT) score decline is that large numbers of lower-income students are now going to college--in large part because of greatly expanded programs of financial aid based upon "need." Since it is well known that test scores are correlated with family income levels, a substantial increase in the percentage of low-income students taking the SAT and going to college would produce a decline in mean scores.

This review of data available failed to find any evidence that would indicate that the score decline is due in any measurable part to a change in the distribution of family income patterns of college-bound or enrolled students after 1972.

College Board Candidate Data

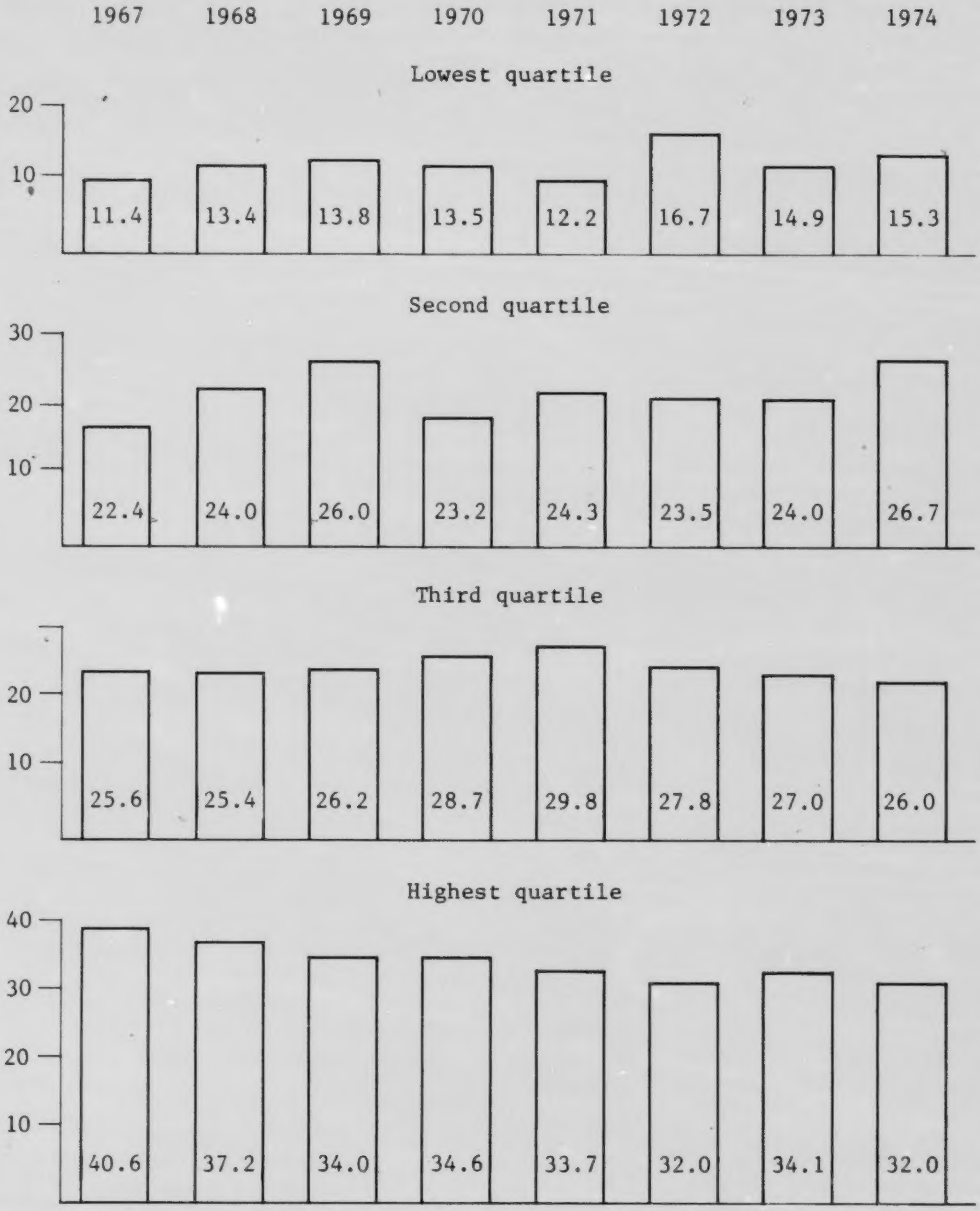
Comparisons of summary Student Descriptive Questionnaire (SDQ) data from the classes of 1971-72 through 1975-76 indicate a rise in reported parental income, as would be expected because of inflation. With inflation considered in constant dollar terms, the percentage of College Board candidates at income levels below $15,000 has remained relatively unchanged over the four-year period for which data are available, and the percentages with incomes above $15,000 has increased slightly. There is no evidence from the SDQ data that would indicate lower-income population reasons for the score decline during the past four years. With no other sources of income-related information available about students who have taken the SAT, a search was conducted for such data that might be descriptive of college-enrolled students over time.

Income Distributions of Entering Freshmen

U.S. Bureau of the Census data indicate that from the mid-1960s to 1972, students from low-income families did indeed enroll in colleges and universities in rapidly rising numbers. Table 1 shows that until 1972, entering freshmen from the lowest fourth of families in terms of income gained ground as a percentage of all entering freshmen, while those from the highest fourth of families declined as a percentage of the total. Young people from the second quartile increased their relative representation slightly while there was also a modest gain for those in the third quartile. This 1967 to 1972 influx of students from lower-income families coincides with the beginnings and ealy years of the score decline, lending credence to the possibility that lower test scores could have resulted from this change in population.

Between 1972 and 1973, however, this trend was reversed and enrollment of low-income young people declined. As illustrated in Table 1, there was an appreciable drop in the relative representation of freshmen in the lowest quartile and an increase in the share of the highest quartile. It might have been expected that test scores would have shown an increase rather than a decrease in 1973 if SAT-takers were representative of the entire freshman population, but this was not the case.

Table 1. Percentage of Entering Freshmen
from Each Family Income Quartile, 1967 to 1974



Source: Carnegie Council on Policy studies in Higher Education, The Federal Role in Postsecondary Education: Unfinished Business, 1975-1980. San Francisco: Jossey-Bass Publishers, 1975, p. 15. 1974 data based on staff estimates of 1974 family income distribution and on freshman family income data from Chronicle of Higher Education, January 20, 1975, p. 8.

Between 1973 and 1974, freshmen from the lowest income quartile regained some lost ground, although their proportion of entering freshmen was still below the 1972 peak. Students from the second quartile also increased their relative proportion, while the percentages from the two upper quartiles fell off. Test scores continued to decline that year, although at a lesser rate. Comparable data are not yet available for 1975 and 1976.

Percent of Age Group in College by Income Levels

Table 2 presents selected college enrollment data for all dependent students, covering the same span of years from 1967 to 1974. It is important to note that although total college enrollment increased dramatically in absolute numbers, the percent enrolled from the total 18-to-24 age group decreased each year from 1969 to 1974. The proportion of black students increased slightly from 1970 to 1974 while the proportion of white students decreased, but the percent of black students enrolled still lags considerably behind that for whites.

With income held in constant dollar terms, the peak of enrollment percentages at various income levels appeared in 1969. At the levels below $10,000, the percentage of the age group enrolled in college shows a small but steady decline each year since 1971, while the percentage above $10,000 remains relatively oonstant.

In 1970, the discrepancies between enrollment rates of blacks and whites based on income were considerable. Among dependent 18-to-24-year-old students in families making less than $10,000, about twice as large a proportion of whites as blacks were enrolled in college. For those with family incomes above $10,000, the difference was smaller. By 1974, however, black and white students in the highest and lowest income ranges were attending in more equal proportion. Clearly, for both groups, the probability of college enrollment tended to increase with income.

As indicated on the bottom line of Table 2, median family incomes increased slightly each year over the eight year span, in constant dollar terms. The lower median income of black families reflects the fact that 64 percent of the 18-to-24-year-old black students who were dependent family members in 1974 were from families with incomes of less than $10,000, whereas 16 percent of white students were from families of the same economic status. One fourth of the black and 43 percent of the white students were from families making $10,000 to $20,000.

Just why the decline in relative proportions of low-income students occurred after 1969 is not clear. In absolute numbers, more low-income students enrolled earh year, but the overall distribution by income level changed only slightly.

Impact of Student Aid on Low-Income Enrollment

Very little comparative information is available on the impact of growing financial aid resources on the characteristics of college-enrolled students. Table 3 presents some comparative data for the years 1969-70

Table 2. College Enrollment Rates for Dependent Family Members by Income Level (Numbers in Thousands: Income in Constant Dollars)

	1974 Total	White	Black	1973	1972	1971	1970 Total	White	Black	1969	1968	1967
Total number enrolled (16 to 34 year olds)	8,589	7,776	813	8,179	8,313	8,087	7,281	6,759	522	7,435	6,960	6,401
Total percent enrolled (18 to 24 age group)	35.5%	37.6	22.2	36.2	37.8	38.4	39.3	42.5	20.4	42.0	40.1	39.1%
Family income levels (of 18 to 24 age group):												
Under $5,000	15.7%	15.6	14.6	15.3	17.3	17.5	17.0	20.6	10.8	19.4	18.5	16.4%
$5,000 to $9,999	26.0	24.5	26.6	26.3	29.1	30.0	33.6	35.1	23.9	32.7	31.2	29.5
$10,000 to $20,000	35.8	36.6	25.7 }	53.7 }	53.7 }	55.1 }	52.0 }	52.8 }	37.1 }	58.5 }	56.5 }	57.1
$20,000 and over	54.2	54.0	54.8 }									
Median family income (18 to 24 years enrolled)	$14,210	$18,500	$9,200	$14,679	$14,190	$13,947	$13,831	$17,800	$8,600	$13,798	$13,638	$13,481

Source: Bureau of the Census, Current Population Reports, P-20, No. 286 and P-23, No. 51.

and 1974-75. While full-time enrollment was growing, and average tuition and board charges were increasing dramatically in real dollar terms over this five-year period, available student aid more than tripled from $1.2 billion to $3.9 billion.

Federal, state, and private aid sources assumed a much larger role in 1975 than was true in 1970. The institutional share of aid support dropped from one-half to one-third. Although private institutions still provide 40 percent of the aid resources available to their students, combined federal and state student aid in 1975 more than equaled the total amount available through private institutions in 1970. The federal aid available to private institutions alone in 1975 exceeds the total federal aid through all institutions in 1970.

Significantly more federal student aid was granted to minority and female students in 1975 than was true for 1970, but in current dollar terms, a much higher percentage (77 percent) of federal student aid in 1970 was provided to low-income students than was the case in 1975 (33 percent). Even if independent, or self-supporting, students are counted as low-income (18 percent), it is evident that inflation, higher college costs, and changed eligibility standards have caused more federal aid dollars to be distributed to middle-income students.

In absolute terms, about the same number of low-income students received federal student aid in 1975 as in 1970, although with higher average award amounts. A significantly larger number of middle-income students were being helped with federal aid by 1975, effectively consuming most of the growth in federal aid since 1970.

Summary

In absolute numbers, universities and colleges enrolled increasing numbers of low-income students over the key years of score decline from 1967 to 1974. At the same time, the percentage of low-income freshmen relative to entering classes increased only until 1972, and the proportions of the college-enrolled age group from low-income families actually declined after 1969. Greatly expanded financial aid programs enabled many more students to enroll in college but appear to have had little impact on changing the overall family income patterns of enrolled students.

There is the possibility that the SAT-taking population of the mid-1960s to 1972 reflected the increased enrollment of university and college freshmen from low-income families, and that some part of the early score decline might be attributed to this shift. There is no evidence to indicate that the score decline after 1972 is due in any measurable part to a change in the distribution of family income patterns of college-bound or enrolled students.

Table 3. Comparison of College Enrollments, Charges, and Characteristics of Aid Available, 1969-70 to 1974-75

	1974-75			1969-70		
	Total	Public	Private	Total	Public	Private
Full-time enrollment (in thousands)	6,617	4,371	2,246	6,080	3,939	2,141
Average charges (tuition, board, room)	-	$1,708	$3,592	-	$1,288	$2,740
*Total aid available (in thousands)	$3,927	$2,027	$1,900	$1,238	$ 495	$ 743
Distribution:						
Federal sources	38.8%	46.9	30.4	32.9%	NA	NA
Institutional:						
Direct	26.0	21.6	30.7	49.2	NA	NA
Waivers	6.8	4.6	9.1			
State and local sources	19.7	18.7	20.7	15.4	NA	NA
Private sources	8.6	8.2	9.1	2.4	NA	NA
Characteristics - all O.E. aid through institutions (in thousands)						
Total	$1,584	$1,033	$ 551	$ 542	NA	NA
Percent	100%	65.2%	34.8%	100%	NA	NA
Characteristics - all O.E.-aided students						
Minority	33.6%	38.3	24.8	28.0%	NA	NA
Nonminority	66.4	61.7	75.2	72.0	NA	NA
Female	51.0	47.7	51.3	41.0	NA	NA
Male	49.0	52.3	48.7	59.0	NA	NA

(continued)

Table 3. (Continued)

	1974-75			1969-70		
	Total	Public	Private	Total	Public	Private
Dependent undergraduate family income:						
Less than $7,500	33.3%	35.6	28.2	77.2%	NA	NA
$7,500 to $11,999	24.8	24.4	25.8	21.8	NA	NA
$12,000 and above	19.1	14.0	30.0	1.0	NA	NA
Independent undergraduates	18.0	22.0	9.3	NA	NA	NA
Graduate students	4.8	3.9	6.8	NA	NA	NA

*Not included are programs outside institutional jurisdiction, e.g., social security, veterans' benefits, Guaranteed Loan Program.

Source: ACE Fact Books, 1975, and ACE 1975 Survey of Student Assistance; and Edward Sanders and James Nelson, "Financing of Undergraduates, 1969-70." Financial Aid News, Vol. 10, No. 2, July 1970.

Available Evidence on Public Attitudes toward Education

John Shelton Reed
Associate Professor, Department of Sociology
University of North Carolina at Chapel Hill

With a supplement of excerpts from
Eighth Annual Survey of the Public's Attitudes toward the Public Schools.
A Project Jointly Conducted by The Gallup Poll and
the Charles F. Kettering Foundation.
Public Opinion Surveys, Princeton, New Jersey, 1976

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

April 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $3 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

This memorandum reports on my search of the archives of American public opinion and survey data for items that could be used to examine (1) changes in public opinion since 1960 that might have affected high school students' Scholastic Aptitude Test (SAT) performance directly (e.g., by providing a "climate" in which academic excellence or college education in general is less highly valued), or (2) changes prior to 1960 that might have affected the quality of their preparation. A list of sources reviewed appears as Appendix A. Although there may be scattered items (or even one or two entire surveys) that have eluded me, I am confident that I have looked at the vast majority of the questions dealing with education asked by the major American survey organizations from 1936 to 1970. My search was less systematic -- and this catalogue is therefore probably least exhaustive -- for the period after 1970, but I assume that changes since then would have little direct bearing on the problems under review by the Advisory Panel on Score Decline, except insofar as one could look at the students of the 1960s as they leave school and enter the pollsters' "population".

Given the quasi-political nature of most of the polls, it is perhaps not surprising that the largest set of questions deals with education as an issue: in particular, with school taxes, federal aid to education, teachers' salaries, and the like. Questions of this sort might be used inferentially to study attitudes toward education, but there are so many extraneous factors affecting responses to them that I excluded them categorically. Similarly, a great many questions dealing with school desegregation, with aid to parochial schools, with high school and college athletics, and other subjects only marginally relevant were not considered further.

With these exclusions, some 130 items remained. For slightly over half of these, frequency distributions were available, either in published sources or because the data are stored in the data library of the Institute for Research in Social Science. These are the data reported in the tables to follow. For the other 60-some items, frequency distributions could be obtained at a modest charge (and the data obtained for analysis at a somewhat less modest charge) from the Roper Center, which holds these data. These items are listed (chronologically, for the most part) in Appendix B. I did not engage the Roper Center to supply me with these data because it is my impression (based on inspection of the data already in hand) that they would be of relatively little use, except perhaps in combination with data from a newly commissioned survey (see below). Similarly, I confined my analysis of the data available here to simple comparison of the frequency distributions across time.

In the data on the following pages, despite the problems introduced by varying question-wordings and missing observations at critical time-points, I believe I discern a few trends -- although whether they are striking enough in magnitude and implication to warrant further investigation is questionable. Let me summarize what I think the pattern is, before dealing with the tables in detail.

After a period of heightened criticism of American education at all levels during the late 1950s, Americans expressed increasing satisfaction with their schools and colleges during the 1960s, until the end of the decade, when they once again came under fire. If this summary is correct (and it is at least generally consistent with the data), it is correlated in an interesting way with the pattern of changes in SAT scores. It is tempting to relate the dissatisfaction of the late 1950s to Sputnik and that of the late 1960s to perceived disorder in the schools and on campuses, but there is of course no way to do this within the framework of these data. (The latter connection is especially doubtful, since Americans' recent disenchantment with education seems to have taken place in the context of a more general disillusionment with our national institutions.) If the critical mood of the late 1950s was followed, as I believe it was, by a period of increasing expressed satisfaction, it is difficult to know whether this was a response to a more satisfactory educational system or simply a reflection of growing indifference.

Table 1 shows the pattern for the 1960s. The observations for 1963 and 1966 are particularly telling, since the questions were worded identically. Allowing for differences in the wording of the questions, there is a fairly steady increase in expressed satisfaction (although the euphoria of 1946 remained unmatched) until the 1970 observation reverses the trend. Table 2 shows the end-of-decade change in greater detail and shows that it applies to colleges as well as to the public schools. "Great confidence" in higher education characterized three out of five Americans in 1967; its absence characterized three out of five in 1971. I should note that nearly every other American institution suffered a concurrent loss of public confidence, and that higher education's relative position -- now second only to medicine -- may even have improved somewhat. But this change is of little interest here, something that happened at the end of the sixties could not have produced a change that got under way at the decade's outset.

Table 3 throws a little more light on the other end of this period. If attitudes toward homework can be taken as a measure of concern for how demanding American high schools are, these data suggest heightened concern around 1958, followed by a gradual decline until at least 1965. The differences between 1954 and 1955 are almost certainly due to the wording of the questions. ("about right" omitted in 1954). But whether 1954 or 1955 is taken as the baseline, there is a substantial increase by 1958 in the proportion of Americans who feel that high school students aren't working hard enough outside school: for the first time, a majority hold this view. By 1961, however, either the situation or people's views of it changed, and by 1965 a solid majority of those with opinions on the matter felt that high school students were doing enough homework. Table 4 shows that, although a majority of Americans felt that the public schools were not demanding enough in 1958, consistent majorities opposed and continued to oppose lengthening the school day or the school year.

Table 5 shows that most Americans in recent years have felt that discipline in the public schools is not strict enough. In this respect, 1970 looks about like the 1950s, and we lack data from the 1960s to see what happened during that interesting decade (although data not reported here show that support for corporal punishment in the schools increased between 1954 and 1958, but declined by the late 1960s). Two surveys from 1954 allow us to say that people seem to feel that the situation in their local schools is better than that in "most schools," and that, in 1954 at least, they are inclined to blame other children's parents rather than school officials for the problem.

Tables 6-9 present an odd lot of items that do not lend themselves to the study of trends but might supply a baseline for a study undertaken today. In particular, the items in Table 7, dealing with how students of different ability should be treated, might shed some light on the rise or decline of meritocratic standards in the United States. Table 9 suggests that such standards were increasingly felt to be appropriate for teachers between 1958 and 1965. Table 8 may be of particular interest to the College Board, although it is only peripherally related to the present topic.

Where do we go from here? If, as I have suggested, obtaining more secondary data for analysis from the Roper Center and analyzing those already in hand more intensively are not likely to be productive, there are two possibilities: (1) to go into the field with a follow-up study in order to obtain trend data based on some of the questions asked earlier, or (2) simply to note that there were some changes taking place in the late 1950s and early 1960s, that they may have had something to do with the decline in test scores, but that it is now too late to study them systematically -- at least with survey data.

If I lean to the latter, it is because I am afraid that data collected now could not be treated convincingly as points on a trend line begun around 1960. Rather, measures now would reflect the general souring of opinion that I believe took place in the late 1960s. I would be reluctant to relate changes in opinion between, say, 1960 and 1976 to changes in SAT scores in the early and middle 1960s.

Table 1. Items Bearing on Satisfaction with Public Schools and personnel

		1946	1948	1955	1963	1965	1966	1967	1970
Are you satisfied with the school your child goes to?[1]	Yes	87%							
	No	13							
On the whole, would you say you are satisfied or dissatisfied with your children's education [education children are getting today]?[2]	Satisfied				66%		75%		
	Dissatisfied				26		17		
	No opinion				8		8		
How would you rate the quality of public school education in this community--[3]	Excellent					26%			
	Pretty good					45			
	Only fair					15			
	Poor					6			
	No opinion					8			
How would you rate the quality of public education in America today--[4]	Excellent								14%
	Pretty good								50
	Only fair								23
	Poor								9
	No opinion								3
Do you think the public schools generally do a good job or a poor job on preparing children for their future?[5]	Good							71%	
	Poor							19	
	No opinion							10	
Do you think children today are being educated better or worse than you were?[6]	Better		59%						
	Same (vol.)		10						
	Worse		26						
	No opinion		5						
[Reading taught better today than when respondent was in school?][7]	Better			36%					
	Same (vol.)			11					
	Worse			41					
	No opinion			12					

(continued)

Table 1 (continued)

		1946	1948	1955	1963	1965	1966	1967	1970
[Teachers in local schools do job--][8]	Well	60%							
	Fair	29							
	Poorly	8							
	No opinion	3							
In general, how would you rate the teachers in the public school system in this community?[9]	Excellent					22%			
	Pretty good					48			
	Only fair					14			
	Poor					2			
	No opinion					15			

1. AIPO 376K. Unclear whether this was asked only of parents. 40 percent had no criticism when asked.
2. AIPO 675K and AIPO 735K. Increase in percent satisfied was 7 percent among whites, 22 percent among blacks.
3. Harris 1522
4. Harris 2050 (women only)
5. AIPO 745A
6. AIPO 420K. Cross-national comparisons available
7. AIPO 552K. College-educated and those over 50 most critical
8. AIPO 376K. See note 1, above
9. Harris 1522

Table 2. Evaluation of College and University Personnel

		1960	1966	1967	1971	1972	1973	1974	1975
Different groups work for one candidate or another. Are there any groups on this list that you particularly trust -- that is, would you be more likely to vote for candidates they recommend? Are there any. . .[for which] you would be more likely to vote against candidates they recommend? -- "College groups"[1]	Trust.	9%							
	No reaction.	90							
	Distrust	2							
As far as the people in charge of running ____________are concerned, would you say you have a great deal of confidence. . .in them? -- higher education[2]	Yes.		61%	62%	37%	33%	44%	40%	36%
	No + no opinion. . .		39	38	63	67	56	60	64

1. SRC Election Study
2. Harris Poll, published reports. Parallel declines were observed for all institutions (except the media) for which this question was asked.

Table 3. Items Related to Homework in the Public Schools

		1954	1955	1958	1961	1965
Do you think high school students today are given too much or not enough homework to do?[1]	Too much.	8%				
	About right (vol.).	26				
	Not enough.	40				
	No opinion.	26				
Do you think students in our high schools should be given more homework, less homework, or do you think they are given about the right amount now?[2]	Give less		6%			
	About right now . .		38			
	Give more		25			
	No opinion.		31			
Should high school students be required to put in more hours of homework each week?[3]	No.			33%		47%
	Yes			51		39
	No opinion.			16		14
Would you favor or oppose increasing the amount of school work done at home by students -- in high schools?[4]	Oppose.				39%	
	Favor				46	
	No opinion.				15	
-- in grade schools?	Oppose.				46%	
	Favor				41	
	No opinion.				13	

1. AIPO 538
2. AIPO 555
3. AIPO 607 and AIPO 710
4. AIPO 642. Published figures differ slightly from these. In each case, these data show 5-6 percent more "don't knows" and correspondingly less opposition.

Table 4. Items Related to Time Spent in School

		1958	1961	1965
Do you think our public schools today demand too much work or too little work from the students?[1]	Too little.	51%		
	Too much.	6		
	About right (vol.).	33		
	No opinion.	10		
Should high school students be required to put in more hours of work in school each week?[2]	Yes	35%		36%
	No.	52		52
	No opinion.	13		12
Would you favor or oppose increasing the number of hours spent in classrooms per day -- in high schools?[3]	Favor		42%	
	Oppose.		50	
	No opinion.		9	
-- in grade schools?	Favor		33%	
	Oppose.		60	
	No opinion.		8	
Would you favor or oppose increasing the number of days per year spent in school -- for high school students?[4]	Favor		28%	
	Oppose.		63	
	No opinion.		9	
-- for grade school students?	Favor		25%	
	Oppose.		67	
	No opinion.		8	
[Opinion of plan to keep schools open year round, with summer devoted to arts and skills][5]	Favor			46%
	Oppose.			47
	No opinion.			6

1. AIPO 595K
2. AIPO 607 and AIPO 710
3. AIPO 642
4. AIPO 642. In all items from this survey, the published reports show somewhat fewer "don't knows."
5. AIPO 710

Table 5. Items Related to Discipline in the Public Schools

		1954	1957	1965	1970
Do you think discipline in the public schools in this community is too strict or not strict enough?[1]	Too strict.	1%			
	About right	38			
	Not strict enough . . .	43			
	No opinion.	18			
Do you think discipline in most schools today is too strict, or not strict enough?[2]	Too strict.	2%			
	About right	25			
	Not strict enough . . .	65			
	No opinion.	8			
(Do you think discipline in most homes today is too strict or not strict enough?)[3]	Too strict.	1%			
	About right	12			
	Not strict enough . . .	83			
	No opinion.	4			
Who do you think is more to blame for [the lack of discipline in the schools] -- the school officials for being too easy on the children, or the parents for not allowing school officials to be stricter?[4]	Parents	41%			
	Officials	6			
	Both, equally	17			
	No lack of discipline .	34			
	No opinion.	1			
Do you think teachers should be permitted to discipline their students more strictly than they do at present, or not?[5]	Should.		70%		
	Should not.		21		
	No opinion.		9		
In general, do you feel young people in elementary school and high school have too much freedom, too little freedom, or about the right amount of freedom?[6]	Too much.				49%
	About right				38
	Too little.				5
	No opinion.				8
Do you think high school students who won't behave and won't do their work should be allowed to stay in high school?[7]	Yes	29%			
	No.	64			
	No opinion.	7			

(continued)

Table 5 (continued)

		1954	1957	1965	1970
[Opinion of strict school dress code][8]	Favor			44%	
	Oppose.			50	
	No opinion.			6	

1. AIPO 533
2. AIPO 538
3. AIPO 538
4. AIPO 538
5. AIPO 587
6. Harris Poll 2050 (women respondents only)
7. AIPO 533
8. AIPO 710

Table 6. Items Related to Public School Curriculum

		1954	1955	1957	1965
Educators do not all agree on how much time should be spent teaching reading in the public schools. Do you, yourself, think reading is more important or less important than other school subjects?[1]	More important. . . .	67%			
	About same (vol.) . .	25			
	Less important. . . .	4			
	No opinion.	3			
If a pupil is having trouble with reading, do you think he should spend extra time catching up, if it means taking time away from other subjects?[2]	Yes	67%			
	Qualified	11			
	No.	17			
	No opinion.	4			
Do you think that grade schools today pay enough attention to teaching pupils how to read?[3]	Yes		39%	38%	
	No.		48	41	
	No opinion.		13	21	
-- how to spell?[4]	Yes			38%	
	No.			41	
	No opinion.			21	
-- how to express themselves in writing?[5]	Yes			38%	
	No.			37	
	No opinion.			25	
[Is there too much emphasis on extracurricular activities?][6]	Yes			38%	
	No.			49	
	No opinion.			13	
It has been suggested that a course in ethics or rules of conduct be given to students in high schools. Do you approve or disapprove of this plan?[7]	Approve				88%
	Disapprove.				5
	No opinion.				6

1. AIPO 533
2. AIPO 533
3. AIPO 552K and AIPO 587K
4. AIPO 587K
5. AIPO 587K
6. AIPO 587K
7. AIPO 710

Table 7. Items Related to Treatment of Students with Differing Ability

		1954	1958	1961
Should every public school pupil in the first eight grades be promoted at the end of the year, or should pupils who have failed most of their subjects be kept back?[1]	Promoted.	13%		
	Kept back	75		
	Qualified	8		
	No Opinion.	5		
Do you think that studies in the public schools should be made easy enough that every child can pass?[1]	Yes	16%		
	No.	78		
	Qualified	2		
	No opinion.	4		
In some public schools, pupils are divided into above-average and below-average sections for study and recitation, according to the scores they make on intelligence tests. Do you think this is a good idea or a poor idea?[1]	Good idea	61%		
	Poor idea	26		
	No opinion.	13		
Some students entering high school can't read any better than the average third grader. Do you think very poor readers should be admitted to high school, or should they be kept back until they have caught up in reading?[1]	Admit	42%		
	Keep back	50		
	No opinion.	7		
Some pupils don't learn much from books after the eighth grade. Our laws require them to stay in school anyway until they reach a certain age. What do you think we should do about high school students who don't like high school and aren't getting anything out of their studies?[1]	Make them go, as now	13%		
	Provide different training.	71		
	Let them stop earlier	10		
	No opinion.	6		
Instead of grouping students according to their scores on intelligence tests, some public schools group them according to their skill in each particular subject. . . .Do you think this is a good idea or a poor idea?[1]	Good idea	71%		
	Poor idea	15		
	No opinion.	15		

(continued)

Table 7 (continued)

		1954	1958	1961
If your child had trouble learning to read, would you prefer that he study reading with other slow readers, or in a class where most of the other pupils read better and faster?[1]	With slow	67%		
	With better	26		
	No opinion.	7		
Should teachers devote extra time to the bright students?[2]	Yes		28%	
	No.		64	
	No opinion.		8	
. . . . to the slow students?[2]	Yes		74%	
	No.		12	
	No opinion.		13	
Some people say that colleges should raise their entrance requirements, making it harder for high school graduates to enter. Do you think colleges should do this, or not?[3]	Should.		27%	25%
	Should not.		62	64
	No opinion.		11	11

1. AIPO 533
2. AIPO 607
3. AIPO 595K and AIPO 642

Table 8. Items Related to Nationwide Testing Programs

		1958	1961	1965
Should all high school students in the United States be required to pass a standard nationwide examination in order to get a high school diploma?[1]	Yes	43%	47%	48%
	No.	36	43	43
	No opinion.	21	9	9
Would you favor or oppose a nationwide test to be given to all high school students when they enter and when they graduate to enable parents to learn more about their children and their educational achievement?[2]	Favor		74%	
	Oppose.		16	
	No opinion.		10	

Table 9. Miscellaneous Items Related to the Teaching Profession

		1957	1958	1961	1965	1968
Would you like to have a daughter of yours go into teaching or not?[3]	Yes.	74%				
	No	16				
	No opinion . .	10				
If you were advising a young person on choosing a career, would you advise this person to become a teacher, or not?[4]	Yes.					62%
	No					24
	No opinion . .					14
Should pay raises for teachers in the public school system be based on teaching ability or on the length of service?[5]	Ability. . . .		66%	70%	74%	
	Service. . . .		17	16	15	
	Both (vol.). .		7	4	4	
	No opinion . .		10	10	7	

1. AIPO 607, AIPO 642, and AIPO 710
2. AIPO 642
3. AIPO 587K
4. AIPO 770D
5. AIPO 607, AIPO 642, and AIPO 710

APPENDIX A. SOURCES SEARCHED FOR SURVEY ITEMS

Hastings, Philip K., and Southwick, Jessie C., eds., Survey Data for Trend Analysis: An Index to Repeated Questions in U.S. National Surveys Held by the Roper Public Opinion Research Center

The Gallup Poll, published reports

The Harris Poll, published reports

Institute for Research in Social Science, Descriptions of Harris Public Opinion Polls Available for Dissemination. Chapel Hill, N.C.: University of North Carolina

Inter-University Consortium for Political Research, Codebooks for Survey Research Center Election Studies, 1952-1972. Ann Arbor, Mich.: University of Michigan

Inter-University Consortium for Political Research, Codebooks for National Opinion Research Center General Social Surveys. Ann Arbor, Mich.: University of Michigan

Roper Center for Public Opinion Research, American Item Index (file of questions from all Gallup Poll surveys, selected Roper and NORC surveys, and miscellaneous others)

APPENDIX B. AVAILABLE ITEMS, FOR WHICH DATA WERE NOT OBTAINED

The questions listed here from the two Roper polls (1945 and 1949) are presented separately because each had a substantial battery of good items that would bear repeating in 1976. The note "Tab only" on the questions for the earlier study means that frequency distributions are available from the Roper Center but not the original data.

Items from Roper Study for Fortune, No. 46, January 1945
(Sample Size 3,500 - 4,499)

Question 2. Tab only.
After the war, if you had a son/daughter graduating from high school, would you prefer that he/she go on to college, or would you rather have him/her do something else, or would you not care one way or the other?

Question 3. Tab only.
A. People have different ideas about what they want their children to get out of college. Which of these things would you rate as very important for a son of yours to get out of college if he were to go and which would you rate as not so important? (List on reverse side.)
B. Now which one or two of these things you feel are very important would you consider most important for a son of yours to get out of college?

A. Training to fit him for a specific occupation or profession
B. The advantage of being recognized as a college man

C. A larger amount of factual information
D. A better appreciation of such things as literature, art, and music
E. Ability to think more logically
F. Ability to get along with and understand people
G. Desire and ability to be a more useful citizen
H. Moral growth

Question 4. Tab only.
A. Do you think there are many or only a few young people who are capable of going through college but do not go because they cannot afford to?

Question 5. Tab only.
In general do you think our colleges and universities are now doing as much as they should on -- ----, or do you think they should be doing more, or less?
A. Training students to be more useful citizens
B. Training students for positions of leadership
C. Training students in specialized fields like labor relations, government services, etc.
D. Setting up projects to help work out important problems facing local communities or the country as a whole
E. Educating the general public through lectures, radio programs, etc.
F. Providing scholars and scientists with opportunities for research

Question 6. Tab only.
Do you think ----- should or should not try to direct the way a professor teaches his subject in a college or university?
A. The other professors
B. The board of trustees
C. The government
D. The people who give large sums of money to a college
E. The people who have graduated from the college

Question 7. Tab only.
If the government should start giving some financial support to those colleges and universities now supported largely by money from private people, do you think the professors would have more freedom in teaching, less freedom, or about the same as now?

Items from Roper Study for Fortune, No 76, April 1949
(Sample Size 5,400 - 6,499)

Question 1.
A. If you had a boy graduating from high school, would you personally like to have him go on to college, or would you rather have him do something else?
B. Would you personally like to have him go to college if he wanted to or would you rather have him do something else?
C. Why would you like to have a son of yours go to college? What would you want him to get out of it?
D. What are some of the reasons why you do not want him to go to college?

Question 2.

A. If you had a girl graduating from high school, would you personally like to have her go on to college, or would you rather have her do something else?

B. Would you personally like to have her go to college if she wanted to or would you rather have her do something else?

C. Why would you like to have a daughter of yours go to college? What would you want her to get out of it?

D. What are some of the reasons why you do not want her to go to college?

Question 3.

A. Which three things on this list would you think were most important of all for college to give your son?

B. Which ones would you think were least important?

A. Desire and ability to be a more useful citizen
B. Preparation for marriage and family life
C. Training for a particular occupation or profession
D. Learning how to get along with all sorts of people
E. A better appreciation of such things as literature, art, and music
F. A broader knowledge of the world and world problems
G. A sharper, better trained mind in dealing with all sorts of problems
H. Development of a good moral character
I. A better chance to get ahead in the world
J: The intelligence and wisdom necessary to live a full life

Question 4.

A. Which three things on this list would you think were most important of all for college to give your daughter?

B. Which ones would you think were least important?

A. Desire and ability to be a more useful citizen
B. Preparation for marriage and family life
C. Training for a particular occupation or profession
D. Learning how to get along with all sorts of people
E. A better appreciation of such things as literature, art, and music
F. A broader knowledge of the world and world problems
G. A sharper, better trained mind in dealing with all sorts of problems
H. Development of a good moral character
I. A better chance to get ahead in the world
J. The intelligence and wisdom necessary to live a full life

Question 5.

In college, are there any subjects you think all boys should be required to study, no matter what they are going to do afterwards? What subjects?

Question 6.

In college, are there any subjects you think all girls should be required to study, no matter what they are going to do afterwards? What subjects?

Question 7.
So far as you are concerned, would you lean toward having a son of yours go to:

A. A college for boys only or to one that has both boy and girl students?
B. A college with a large number of students/more than a thousand/ or to a small college?
C. A college that has the kind of social clubs called fraternities or to a college that does not have them?

Question 8.
So far as you are concerned, would you lean toward having a daughter of yours go to:

A. A college for girls only or to one that has both boy and girl students?
B. A college with a large number of students/more than a thousand/ or to a small college?
C. A college that has the kind of social clubs called sororities or to a college that does not have them?

Question 9.

A. Which would you like a son of yours to do -- spend all his time in college on general, liberal arts education, all of it on professional education, or some of it on both?
B. About what share of his time in college would you like to see him spend on general, liberal arts education -- a quarter, a half, or three quarters?

Question 10.

A. Which would you like a daughter of yours to do -- spend all her time in college on general, liberal arts education, all of it on professional education, or some of it on both?
B. About what share of her time in college would you like to see her spend on general, liberal arts education -- a quarter, a half, or three quarters?

Question 13.
If you had to pick one college or university in the United States where a boy could get the best education, regardless of cost, what one would you pick?

Question 14.
If you had to pick one college or university in the United States where a girl could get the best education, regardless of cost, what one would you pick?

Question 15.

A. It has been proposed that the federal government in Washington provide money to send qualified high school students to college who otherwise could not afford to go. I would like you to read these arguments and tell me whether you are in favor of, or against, such a plan.
B. Can you tell me which one of the three arguments for/against the plan had the most influence on your decision?

Arguments in favor of plan:

A. There are lots of qualified high school graduates who cannot afford to go to college but who should have the chance.

B. Colleges need the extra money, since they cannot rely on gifts from rich men as much as they used to do.

C. The federal government would help out the students and colleges in the parts of the country that need help most.

Arguments against plan:

A. The federal government cannot afford it -- it is spending too much money now.

B. In the long run it would mean that the federal government would have too much influence on what the colleges teach.

C. If any part of government is going to pay for student expenses it should be the state government, not the federal government.

Question 16.

Here is a list of words, and I would like you to look them over and pick the three or four that best describe your idea of what the average college student is actually like: immoral, hard-working, radical, intelligent, well-informed, soft, time-wasting, well-mannered, overworked, conservative, hard-drinking, snobbish, ambitious, democratic.

Question 17.

Should colleges have classes that take up the subject of ---- or should ----- be discussed in classes only when students ask about it, or would it be better not to discuss it at all?

A. Racial and religious prejudice

B. Sex education

C. How to be a good parent

D. Religious beliefs

E. Communism

Question 18.

Here are several different ways in which college can deal with the subject of socialism and capitalism in class discussions with students. Which do you think is best?

1. Allow the professors to give only the arguments in favor of capitalism to students.
2. Require any professor who discusses the subject to give the arguments for both sides, but let him express his own opinion only if it is favorable to capitalism.
3. Require any professor to give the arguments for both sides, then let him express his own opinion, whatever it is, but try to have professors with differing opinions on the faculty.
4. Require any professor to give the arguments for both sides, but do not allow him to express any opinion of his own.
5. Do not know what socialism and/or capitalism is.

Question 19.

A. As you know, businessmen and rich people help support private colleges with gifts. Do you think these wealthy people ever try to influence what these colleges teach students about capitalism and socialism, or not?

B. Would you say these wealthy people actually do have a good deal, a little, or no influence at all on what these colleges teach students about capitalism and socialism?

Question 20.
Do you think a man who gives a large sum of money to a college should or should not have a right to some say on what the college teaches its students about capitalism and socialism?

Question 23.
Some experts believe the following statements are true, while other experts believe they are false. Would you say this statement was more true than false, or more false than true?
A. You have got to have a college education these days if you want to get ahead in the world.
B. Most college graduates cannot even write a good letter.
C. Most college professors are not practical enough to run a business successfully.
D. A college education is just as important for a girl who wants to get married right after graduation as for a girl who wants to have a career.
E. There is a lot of communism being taught in colleges today.
F. The average student in a private college pays for about half of what his education actually costs and then the college pays the rest.
G. It would be a good idea for all college students to work for at least part of their college expenses.
H. Since there are business and professional men on most boards of trustees that run colleges, there ought to be labor union officials who are trustees also.
I. If a college turns out first-rate football teams, it is a pretty good sign that it will also provide its students with a good all-round education.
J. There are lots of young people in our colleges now that should not be there.
K. There are lots of young people not now in college who should be there.
L. A jewish student has a harder time getting into most colleges than a non-Jewish student with the same high school marks.

Question 27.
B. Do you think the following statement is more true than false, or more false than true: A boy/girl fresh out of college is not much use to an employer until he/she has unlearned a lot of the things the professors taught him/her.

Miscellaneous Other Items

AIPO: American Institute of Public Opinion (Gallup)
RFOR and RCOM: Roper Associates
NORC: National Opinion Research Center
OPOR: Office of Public Opinion Research

Year	
1936	7. Do you think a college education is worth what it costs in time and money to persons who are not going into professions? (AIPO 59)
1938	2. Which do you think has a better chance of earning a living today -- a high school graduate who has had four years of experience, or a man just out of college? (RFOR 3)
1939	6a. Do you feel you have had enough education? (AIPO 150)
	16. Have you sent or do you intend to send your children to college? Why or why not? (RFOR 13)
1940	2a. Some successful people think too much importance is placed on high school and college education. Do you agree with this? (AIPO 183)
	1. Many people say there is too much importance placed on education these days. Do you agree or disagree? Why? (OPOR 301)
	2a. Do you think young people today are getting a better education in school than their parents got? Why? (OPOR 301)
	[Are some high school students] better off at work? (OPOR 301)
1943	15a. How far in school do you think young people should be required to go today? (AIPO 297K)
	15a. [preamble] For most people, do you feel education is absolutely necessary to be a success in life? (AIPO 298)
	12. [What would you say is] the most important thing for children to get from education? (NORC 217 -- repeated in 1955)
	13a. Are you satisfied with what [your children] are getting from their education in school? (NORC 217 -- repeated in 1955)
1945	Large battery of items (RFOR 46 -- see Appendix B.1)
	3. . . .Around here, would you say that the schools are doing an excellent, good, fair, or poor job? (NORC 238 -- repeated in 1947)
1947	Same as Question 3 on NORC 238 (1945). (NORC 245)
1948	18. Generally speaking, do you think a person has to have a college education if he is going to get and keep a really good job, or do you think it just makes it easier for him, or do you think he can do as well or better by spending those years on something else? (RFOR 71)
	6. In general, would you say you are reasonably well satisfied with -- the kind of education young people are getting today? (AIPO 434T)
1949	Large battery of items (RFOR 76 -- see Appendix B.1)
	Do you think that grade schools today pay enough attention to teaching students how to spell or not? (AIPO 447 -- repeated in 1957 [see text])
1950	1. In general, would you say that school children today are being taught more worthwile and useful things than children were twenty years ago, not as [many], or about as [many] as then? (RCOM 40)

2. [Would you say you are] very satisfied, only fairly well satisfied, or not very satisfied with the public school system in your community? (RCOM 40)

3a. Up to what age do you think the law should require that children go to school? (RCOM 40)

7. What do you think are the two or three most important things young people should get out of high school? (RCOM 40)

9. Compared to twenty years ago, [do the schools have] better trained and more capable teachers. . . ? (RCOM 40)

1952 12a. If you had the time would you like to spend a few hours each week studying a subject such as those taught in college? Which of these subjects would you most like to study now? (AIPO 484)

1955 22. [What would you say is] the most important thing for children to get from education? (NORC 370 -- also asked in 1943)

Are you satisfied with what [your children] are getting from their education in school? (NORC 370 -- also asked in 1943)

24. [Are there] any changes you'd like to see in the public schools? (NORC 370)

1958 2. [preamble] What is the biggest thing wrong with our schools? (AIPO 595)

1959 Would you favor increasing the amount [of homework] in high school? In grade school? (AIPO 613 -- repeated in 1961 and 1965 [see text])

1960 Respondents asked to identify "problems in this community" (from list) -- poor quality of education (AIPO 634)

1965 On the whole, would you say that you are satisfied or dissatisfied with your children's education? (AIPO 715 -- also asked in 1963 and 1966 [see text], again in 1966 and twice in 1969)

1966 Same as items from AIPO 715 (1965). (AIPO 734)

1968 In general, how much respect would you say you have for today's college students as a whole -- a great deal, some, or hardly any? (AIPO 760)

1969 Same as item from AIPO 715 (1965). (AIPO 777 and AIPO 778)

Excerpts from Eighth Annual Survey of
the Public's Attitudes
toward the public schools.
A Project Jointly Conducted by
The Gallup Poll and the
Charles F. Kettering Foundation.

Public Opinion Surveys, Princeton, New Jersey, 1976.

PURPOSE OF THE STUDY

The eighth annual survey of the series has attempted to measure the attitudes of Americans toward their public schools. Each year great care is taken to include new issues of concern to both educators and the public, as well as trend questions that have ongoing impact in the educational world.

This year's survey has been funded by /I/D/E/A, Institute for Development of Educational Activities, Inc., an affiliate of the Charles F. Kettering Foundation. In prior years, these studies have been sponsored by CFK, Ltd. and The Ford Foundation.

To make certain that the survey would embrace important issues in the field of education, a group of educators, along with the project directors from Gallup, met in Miami to discuss and evaluate preliminary questions and possible issues to be included in this study.

Panel members included:
James Betchkal, Editor, *American School Board Journal*, National School Boards Association, Evanston, Illinois
Edward Brainard, Director, Leadership Development, Colorado Department of Education, Denver, Colorado (Now Professor of Education, University of Northern Colorado, Greeley, Colorado)
B. Frank Brown, Division Director, /I/D/E/A, Melbourne, Florida
Stanley M. Elam, Editor, *Phi Delta Kappan*, Bloomington, Indiana
William J. Ellena, Superintendent, Charlottesville Public Schools, Charlottesville, Virginia
Sidney P. Marland, Jr., President, College Entrance Examination Board, New York, New York
William Mitchell, Superintendent, Greeley Public Schools, Greeley, Colorado
John W. Porter, State Superintendent of Public Instruction, Department of Education, Lansing, Michigan
Samuel G. Sava, Executive Director, /I/D/E/A, and Vice President for Educational Activities, Charles F. Kettering Foundation, Dayton, Ohio
Suzanne Sepper, Teacher, Bayview Elementary School, Fort Lauderdale, Florida
M. Donald Thomas, Superintendent of Schools, Salt Lake City School District, Salt Lake City, Utah

We wish to thank these panel members for their valuable suggestions.

George Gallup

RESEARCH PROCEDURE

The Sample

The sample in this eighth annual survey included a total of 1549 adults (18 years and older). It is described as a modified probability sample of the nation. Personal, in-home interviewing was conducted in every area of the nation and in all types of communities. A full analysis of the sample is in the appendix to this report.

Time of Interviewing

The fieldwork for this study was conducted during the period April 21 through April 25, 1976.

The Interviewing Form

Questions included in the survey instrument were selected following extensive pretesting by the trained staff of interviewers maintained by the Gallup organizations.

Findings from this report apply only to the country as a whole and not to specific local communities. Local surveys, using the same questions, can be conducted to determine how local communities compare with the national norm.*

INTRODUCTION

Aside from its historical significance, the year 1976 may well prove to be the turning point in the public's attitudes toward the public schools. Evidence from the present survey indicates a levelling off in the downward trend of recent years in the public's attitudes toward the public schools.

The public schools, as other public institutions, reflect the major trends in society. And the year 1976 has witnessed a nationwide shift toward more traditional values in almost every field.

Whether we have come to the end of the era of permissiveness is yet to be seen. The fact remains, however, that the public is now demanding stricter rules dealing with the behavior of the young and higher standards in the public schools.

Juvenile delinquency, increasing yearly, has focused attention on the need for moral education not only in the home but in the schools. Laws in many states are being changed to permit stiffer sentences for young offenders. It is hoped that such laws and new ways of dealing with the problem of discipline that are being tried in schools, will bring about a lessening of the turmoil found in so many cities.

A growing demand to place greater emphasis on the basics in the school curriculum is evidenced in the findings of the present survey. Meeting this demand could result in higher national test scores and increased respect for the public schools.

---

*Suggestions for such a local survey are outlined in The Gallup Polls of Attitudes toward Education (1969-1973), Chapter 7, "A look into Your School District." Available from Phi Delta Kappa, 8th and Union Streets, Bloomington, Indiana 47401. $2.25.

MAJOR PROBLEMS CONFRONTING THE PUBLIC SCHOOLS IN 1976

Discipline continues to head the list of major problems when a sample of the nation's adults is asked to cite what they perceive as the most important problems of the public schools in their own communities. In fact, discipline has been named most often seven times during the last eight years.

The percentage of respondents citing discipline as the number one problem, however, has shown no increase during recent years. Next to discipline, and in second place, are the problems associated with busing and integration. In third place this year is lack of proper financial support. These were also named second and third, respectively, in last year's survey.

The one significant change from 1975 is the marked increase in the number of persons in the sample who cite "a poor curriculum." This complaint, listed seventh in importance last year, has moved up to fourth place, undoubtedly because of the wide publicity given to the drop in national test scores and growing concern about the number of functional illiterates among school leavers.

Below, in order of mention, is the list of major problems of the local public schools as perceived by the residents of these communities:

1. Lack of discipline
2. Integration/segregation/busing
3. Lack of proper financial support
4. Poor curriculum
5. Use of drugs
6. Difficulty of getting "good" teachers
7. Parents' lack of interest
8. Size of school/classes
9. School board policies
10. Pupils' lack of interest

1976 RATING OF THE PUBLIC SCHOOLS

A five-point scale, familiar to the American public, was employed for the first time in 1974 to establish a base for measuring the public's perceptions of the quality of public school education in their own communities.

The question asked is this:
"Students are often given the grades A B C D and Fail to denote the quality of their work. Suppose the public schools themselves, in this community, were graded in the same way. What grade would you give the public schools here -- A B C D or Fail?"

A significant drop in the public's ratings was recorded between the years 1974 and 1975. However, the decline has now levelled off, with ratings for 1976 being virtually the same as for 1975.

Here are the ratings given the public schools nationally for the last three years:

Ratings given the public schools	National totals 1976 percent	 1975 percent	 1974 percent
A rating	13	13	18
B rating	29	30	30
C rating	28	28	21
D rating	10	9	6
Fail	6	7	5
Don't know/no answer	14	13	20

If the top two positions -- A and B -- are combined, and the bottom three -- C D and Fail -- are combined, the findings indicate that the lowest ratings come from those parents who send their children to private or parochial schools. Those with children in the public schools give the highest ratings to the schools, but even in the case of this group only 50 percent give the schools an A or B rating. This contrasts with only 34 percent of those whose children attend private or parochial schools who give an A or B rating.

	National totals percent	No children in schools percent	Public school parents percent	Parochial school parents percent
A rating	13	12	16	6
B rating	29	26	34	28
C rating	28	26	30	40
D rating	10	9	10	12
Fail	6	7	5	4
Don't know/no answer	14	20	5	10

When the ratings given by the major socioeconomic groups are analyzed, it is evident that the people who are least satisfied with the quality of education offered in their local schools are the recent graduates, the 18 to 29 age group, those who live in the center cities as opposed to the suburbs, and citizens who live in the Western states.

The Question: "Students are often given the grades A B C D and Fail to denote the quality of their work. Suppose the public schools themselves, in this community, were graded in the same way. What grade would you give the public schools here -- A B C D or Fail?"

	A percent	B percent	C percent	D percent	Fail percent	Don't know/ no answer percent
National.	13	29	28	10	6	14
Sex:						
Men.	11	32	28	11	4	14
Women.	14	27	29	9	7	14
Race:						
White.	13	30	28	9	6	14
Nonwhite	10	21	33	13	9	14
Age:						
18 to 29 years	8	25	33	15	6	13
30 to 49 years	15	30	30	10	5	10
50 years and over.	15	31	22	6	7	19
Education:						
Grade school	18	22	27	8	5	20
High school.	12	27	30	11	7	13
College.	11	37	25	9	6	12
Community size:						
1 million and over	11	27	32	13	6	11
500,000 - 999,999.	11	34	21	13	7	14
50,000 - 499,999	10	26	32	11	10	11
2,500 - 49,999	13	33	25	8	2	19
Under 2,500.	17	29	28	6	4	16
Central city	7	23	31	16	11	12
Suburbs.	14	34	28	7	5	12
All other.	16	30	27	7	3	17
Region:						
East	11	31	29	10	5	14
Midwest.	16	32	27	8	4	13
South.	13	27	29	9	7	15
West	10	24	30	15	8	13

HOW TO IMPROVE THE QUALITY OF THE PUBLIC SCHOOLS

After having obtained the rating of their school by the different groups in the public, the next question sought to determine what, in the public's opinion, could be done to improve the quality of local education. As an aid, each respondent was handed a card listing a dozen suggestions and had the opportunity to choose as many as he or she wished.

Answers to this question correspond closely with those given in answer to the question asking about the most important problems faced by the local schools.

This approach sheds further light on the public's current concern that the schools should "devote more attention to teaching of basic skills." This is the most popular suggestion, with 51 percent of all respondents choosing this as a way to improve the quality of education in the local schools. Close behind, in fact, with virtually the same number of choices, is the suggestion to "enforce stricter discipline."

The proposals that get the third and fourth highest number of mentions are to "meet individual needs of students" with 42 percent, and to "improve parent-school relations" with 41 percent.

The top three choices of parents whose children now attend the public schools are the same as the top three of the general public. Oddly enough, parents place even higher on their list, in fourth place, "instruction in morals" which has been generally regarded as the sole responsibility of home and church.

The Question: "Which of these ways do you think would do most to improve the quality of public school education overall?"

	National totals percent	No children in schools percent	Public school parents percent	Parochial school parents percent
Devote more attention to teaching of basic skills.	51	47	55	60
Enforce stricter discipline	50	47	52	64
Meet individual needs of students.	42	39	47	44
Improve parent-school relations	41	43	36	47
Emphasize moral development	39	34	45	49
Emphasize career education and development of salable skills.	38	39	36	37
Provide opportunities for teachers to keep up to date regarding new methods	29	27	32	29
Raise academic standards.	27	28	23	38
Raise teachers' salaries.	14	15	16	8
Increase amount of homework.	14	12	17	21
Build new buildings	9	8	12	7
Lower age for compulsory attendance.	5	4	6	1
None.	1	1	*	-
Don't know/no answer.	4	4	2	3

*Less than 1 percent
(Totals add to more than 100 percent because of multiple answers.)

NATIONWIDE TEST FOR GRADUATION

A consensus is found in favor of requiring high school students to pass a standard examination in order to receive a high school diploma.

Present survey findings on this question show a substantial shift in the public's attitude during the last eighteen years. The question posed then by The Gallup Poll is the same one as that used in the 1976 survey.

The question first asked in 1958 was:
"Should all high school students in the United States be required to pass a standard nationwide examination in order to get a high school diploma?"

The findings for the two surveys show:

	National test for graduation	
	1976 percent	1958 percent
In favor of such a test.	65	50
Opposed.	31	39
No opinion	4	11

It is the least well educated, strangely enough, who are most in favor of such a requirement. The same finding was reported in the 1958 study. Those who have had the advantage of one or more years of college education are much more evenly divided: 53 percent are in favor, 44 percent opposed. Those with a grade school education vote 76 percent in favor, 18 percent opposed.

Persons who are in the 18 to 29 year age bracket are less enthusiastic than older persons, possibly because some may believe that such a test would have deprived them of their own diplomas.

Apart from these differences, the survey findings show that a majority favors such a requirement in all sections of the nation and in the large cities as well as the small towns.

The Question: "Should all high school students in the United States be required to pass a standard nationwide examination in order to get a high school diploma?"

	Yes, they should percent	No, they should not percent	Don't know/ no answer percent
National	65	31	4
Sex:			
Men	66	31	3
Women	65	30	5
Race:			
White	65	31	4
Nonwhite.	67	27	6

(continued)

	Yes, they should percent	No, they should not percent	Don't know/ no answer percent
Age:			
18 to 29 years	56	41	3
30 to 49 years	67	30	3
50 years and over	71	23	6
Education:			
Grade school	76	18	6
High school	69	27	4
College	53	44	3
Community size:			
1 million and over	70	25	5
500,000 - 999,999	69	28	3
50,000 - 499,999	67	31	2
2,500 - 49,999	68	31	1
Under 2,500	58	35	7
Region:			
East	65	31	4
Midwest	66	30	4
South	64	31	5
West	67	31	2

THE PUBLIC'S CONCLUSIONS FROM THE DECLINE IN NATIONAL TEST SCORES

The American people, judging from present and past survey data, do not share the same skepticism about standardized tests found among many educators.

They believe, for example, that the decline in national test scores in recent years means that the quality of education today is declining. Of course, the test scores may only confirm what many already think about the schools.

There is surprising unanimity among all groups in the population that the tests are actually measuring a decline in the quality of education. By majorities of about two-to-one, all segments of the population, in all areas of the nation, believe that the tests are correctly assessing the situation.

The first question asked was:

"Do you believe that a decline in national test scores of students in recent years means that the quality of education today is declining?"

By socioeconomic groups:

	Quality of education		
	Yes, is declining percent	No, is not percent	Don't know/ no answer percent
National.	59	31	10
Sex:			
Men.	61	31	8
Women.	58	31	11
Race:			
White.	60	31	9
nonwhite	58	29	13
Age:			
18 to 29 years	52	40	8
30 to 49 years	62	29	9
50 years and over.	63	26	11
Education:			
Grade school	54	25	21
High school.	61	29	10
College.	58	39	3
Community size:			
1 million and over	67	28	5
500,000 - 999,999.	64	29	7
50,000 - 499,999	62	29	9
2,500 - 49,999	58	33	9
Under 2,500.	50	34	16
Region:			
East	62	26	12
Midwest.	56	38	6
South.	56	31	13
West	66	27	7

All respondents were then asked to explain what they thought brought about the decline in test scores. Each was handed a card with eight reasons listed, including one for the doubters, reading "The tests are not reliable."

The second question put to respondents was: "Here are some reasons that have been given to explain the decline in national test scores. Will you look over these reasons and then tell me which ones you think are most responsible for this decline."

Respondents, who were permitted to choose more than one reason for the decline in test scores, voted in this manner:

	National totals percent	No children in schools percent	Public school parents percent	Parochial school parents percent
1. Less parent attention, concern and supervision of the child.	65	64	65	72
2. Students aren't as motivated to do well. .	52	50	57	53
3. Too much television viewing. . .	49	48	51	51
4. Society is becoming too permissive.	49	47	49	61
5. Teachers are giving less attention to students.	39	39	41	32
6. It's easier to get into college now	16	20	10	14
7. Schools are expanding the number of courses offered	10	12	8	4
8. The tests are not reliable.	16	15	16	16
Other and no opinion . . .	14	13	15	15

Overall, the public places the greatest blame for declining test scores on parents, on society, on children's lack of motivation, and on too much television viewing. Interestingly, parents of children now attending school most often cite the very same top reason for declining scores -- "less parent attention, concern and supervision." Parents themselves are readily accepting the blame and are not trying to place the burden on the schools, as one might expect them to do.

School board members and teachers often complain about the lack of citizen interest in school matters. Lack of knowledge about the problems of the schools, especially the financial problems, all too often results in the defeat of school budgets and school bonds.

Both school and community gain when citizens do take an active interest in the schools, but without a way to engage the interest, and to involve citizens in school problems, nothing much happens. A few individuals will attend school board meetings. The fact remains, however, that those people in the community who could make the greatest contribution to education have almost no way to participate in a meaningful way.

School boards, typically, are too busy dealing with financial and staff problems to spend much time on the many educational matters that must be dealt with today. The appointment, by the school board, of citizen advisory committees offers a practical way -- and judging from the survey results, a popular way -- to enlist the interest of an amazing number of persons in the community.

That this plan, now followed in some communities, offers great opportunities to involve citizens in the problems of the schools is fully evidenced by the results of the present survey.

The question asked was:

"Some school boards have appointed citizen advisory committees to deal with a number of school problems such as discipline, the curriculum, textbook selection, teacher evaluation, the athletic program and the like. The committees report their findings to the school board for possible action. If such a plan were adopted here (or exists here), which of these problems would you most like to deal with on a citizen advisory committee?"

Only 10 percent of all those questioned said they would not like to serve on such a committee or could not make up their minds. While many of the other 90 percent who chose committees on which they would like to serve might find it impossible or impractical to carry out their intention, still an extraordinary number of persons in a typical community are sufficiently interested to fill places on a score of committees. And, if such committees were appointed, then hundreds of citizens might be involved in local school matters.

Each person interviewed in the survey was handed a card that listed 20 committees. He was asked if he would like to serve on any one; in fact, he could choose as many as he wished.

The following list shows the percentage of respondents selecting each committee as one on which he would like to serve.

Advisory committees	Percent who would like to serve on such a committee* percent
1. Discipline and related problems	47
2. Student-teacher relations	31
3. Career education	29
4. Student dropouts	29
5. Teacher evaluation	28
6. The handicapped student	26
7. Educational costs-finances	22
8. The curriculum	21
9. Education for citizenship	19
10. Work-study programs	19
11. Home study and work habits	19
12. Community use of school buildings	16
13. Pupil assessment and test results	15
14. School facilities	14
15. Public relations of schools	13
16. School transportation	12
17. The athletic program	12
18. Educational innovations	12
19. Extracurricular activities	11
20. Progress of recent graduates	9
None	4
Don't know/no answer	6

*Adds to more than 100 percent because of multiple answers.

QUALITIES MOST IMPORTANT IN DEVELOPMENT OF A CHILD

In a society that constantly changes, the qualities that should be developed in its children will change also. What qualities do Americans today regard as most important? Which should be given the most attention by parents and schools? Which are the most neglected?

To shed light on these concerns, the following question was included in the survey:

"Of course all of the qualities listed on this card are important in the overall development of a child. But which one do you regard as the MOST important?"

	National totals percent	No children in schools percent	Public school parents percent	Parochial school parents percent
1. Learning to think for oneself.	26	26	26	23
2. Ability to get along with others.	23	25	20	20
3. Willingness to accept responsibility	21	19	21	26
4. High moral standards . .	13	11	15	20
5. Eagerness to learn . . .	11	10	13	10
6. Desire to excel.	4	6	3	1
Don't know/no answer. . . .	2	3	2	-

All respondents were offered the opportunity to make a second choice among these six qualities. When the second choices are added to the first choices, then "Willingness to accept responsibility" moves up to second place, displacing "Ability to get along with others," which drops to third place.

Looking at the same list of qualities, respondents were asked: "Which one do you think is most neglected by parents today?"

	National totals percent	No children in schools percent	Public school parents percent	Parochial school parents percent
1. High moral standards. .	32	29	36	44
2. Willingness to accept responsibility.	30	33	29	23
3. Learning to think for oneself	14	15	11	14
4. Ability to get along with others	8	7	9	7
5. Eagerness to learn. . .	5	4	5	2
6. Desire to excel	4	3	5	5
Don't know/no answer . . .	7	9	5	5

And as a final question in this series:
"And which one do you think is most neglected by schools?"

	National totals percent	No children in schools percent	Public school parents percent	Parochial school parents percent
1. High moral standards .	26	24	28	26
2. Learning to think for oneself.	15	15	14	11
3. Eagerness to learn . .	13	14	12	15
4. Willingness to accept responsibility	12	13	10	20
5. Desire to excel. . . .	9	8	12	11
6. Ability to get along with others.	8	9	8	5
Don't know/no answer. . .	17	17	16	12

HOW THE PUBLIC FEELS ABOUT UNIONIZATION OF TEACHERS

Whether teachers should join unions has been a highly controversial issue in many areas of the United States for many years. At the present time, most teachers of the nation belong to unions or associations that bargain over salaries and working conditions. The question arises as to how the American public presently appraises this situation. Has unionization helped or hurt public school education? This is the question that was put to the public and the replies reveal a mixed reaction.

The question:

"Most teachers in the nation now belong to unions or associations that bargain over salaries, working conditions and the like. Has unionization, in your opinion, helped, hurt, or made no difference in the quality of public school education in the United States?"

	National totals percent	No children in schools percent	Public school parents percent	Parochial school parents percent
Helped.	22	22	23	24
Hurt.	38	38	36	47
Made no difference. . . .	27	26	28	25
Don't know/no answer. . .	13	14	13	4

Older persons are much more likely to say that unionization has hurt the quality of education. The age group 18 to 29 is almost evenly divided. In the bigger cities of the nation, where the unions tend to be more militant, a slightly greater number say that unionization has "hurt."

Here are the findings by socioeconomic groups:

	Helped percent	Hurt percent	Made no difference percent	Don't know/ no answer percent
National	22	38	27	13
Sex:				
Men	24	42	25	9
Women	21	34	29	16
Race:				
White	21	39	28	12
Nonwhite	26	31	22	21
Age:				
18 to 29 years	26	30	34	10
30 to 49 years	24	37	26	13
50 years and over	16	46	23	15
Education:				
Grade school	18	39	25	18
High school	20	37	30	13
College	29	40	23	8
Community size:				
1 million and over	17	40	34	9
500,000 - 999,999	27	38	26	9
50,000 - 499,999	25	39	27	9
2,500 - 49,999	22	37	26	15
Under 2,500	20	36	24	20
Region:				
East	20	43	28	9
Midwest	19	40	29	12
South	27	33	23	17
West	23	37	28	12

RESPONSIBILITY FOR MORAL BEHAVIOR

The American people have reached the conclusion that many parents either won't or can't control the behavior of their children. They are ready, therefore, to turn over part of the responsibility to the schools. In the 1975 survey of The Public's Attitudes Toward the Public Schools it was reported that 79 percent of all those questioned favored instruction in morals and moral behavior in the schools. Only 15 percent opposed. It was also pointed out in the same report that parents of children now in school were, of all groups, most in favor of this instruction.

In the present survey, a sizable majority of those questioned want part of the responsibility for moral behavior turned over to the schools. Just how this can be done raises many issues. One hundred years ago, McGuffey's Readers pointed out a moral in almost every paragraph. And thousands of years ago, parables performed this service. The modern equivalent may be the case history, as discussed in the 1975 report.

The Question: "Parents now have responsibility for the moral behavior of their children. Do you think that the schools should take on a share of this responsibility, or not?"

	National totals percent	No children in schools percent	Public school parents percent	Parochial school parents percent
Yes, schools should take on share of responsibility.	67	65	69	69
No, they should not . . .	30	31	29	26
Don't know/no answer. . .	3	4	2	5

The Question: "Do you believe that a decline in national test scores of students in recent years means that the quality of education today is declining?"

	National totals percent	No children in schools percent	Public school parents percent	Parochial school parents percent
Yes, is declining. . . .	59	60	58	64
No, is not	31	30	32	28
Don't know/no answer . .	10	10	10	8

Annotated Bibliography

Elsa J. Rosenthal
Albert E. Beaton
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

May 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

American Council on Education, "National Norms for Entering College Freshmen -- Fall 1966." ACE Research Reports, Vol. 2, No. 1, 1967.
American Council on Education, "The American Freshman: National Norms for Fall 1972." ACE Research Reports, Vol. 7, No. 5, 1972.

The ACE has developed norms for entering college freshmen annually for many years and for many variables, including racial composition. For the freshman year beginning in the fall of 1966, 10.8 percent in public four-year colleges were either black or American Indian (ACE 1967, p. 22). In 1972, this figure had increased to 15.9% (ACE 1972, p. 33). In 1972, 2.0 percent of entering college freshmen in public four-year colleges reported that they were either Mexican American or Puerto Rican (these categories had not been included in the 1966 questionnaire). That such increases in minority enrollment were not limited to public institutions is also indicated. In 1966, 2.4 percent of students in private universities were black or American Indian. In 1972, this figure was 5.4 percent.

Angoff, W., "Why the SAT Scores are Going Down." English Journal, March 1975, pp. 10-11.

Angoff reviews some hypotheses about factors that may have contributed to the decline of SAT scores. He also discusses others that appear not to have had any influence on the drop in scores. The hypotheses fall into four categories: (1) that there has been a change in the abilities of the high school population; (2) that there has been a change in the college-going population; (3) that there has been a change in the population that takes the SAT and similar college-entrance tests; (4) that the change is an artificial result of test properties and testing techniques.

Armbruster, Frank, The U.S. Primary and Secondary Education Process. With contributions by Paul J. Bracken and the assistance of Joyce Lind. Final Draft. Croton-on-Hudson, N.Y.: Hudson Institute, 1975.

This study was conducted as part of a project of United States social policy studies conducted under a grant from the U.S. Office of Economic Opportunity. Particularly relevant to the concerns of the Advisory Panel on Score Decline is an appendix citing test results from 22 annual state testing programs, from the District of Columbia and several large cities. The test scores show that the growth in academic achievement noted after World War II topped out during the 1960s. Many achievement indicators show this turning point occurring after the middle of the 1960s.

In a substudy by Paul J. Bracken, data are also cited from national renorming studies that point to significant declines in scores, especially in the upper grades. The Iowa Tests of Basic Skills were renormed in 1955, 1964, and 1970; the Comprehensive Tests of Basic Skills in 1973; and the Iowa Tests of Educational Development (ITED) in 1971. ITED reports also indicate that scores have continued to fall since 1971.

Armbruster rules out socioeconomic factors as major contributors to lower aptitude and achievement and concludes: "the problems of education are more likely to originate from forces within the schools than from outside factors. . . Perhaps a contributing factor was the same one which may have caused academic demands to be reduced: a loss of perspective by educators." (p.275) With the loss of perspective has come an erosion of traditional values and a decline in both social and academic discipline. Armbruster suggests that "our implicit trust in educators may not be well founded. . . Perhaps the most important first step is to convince those with the final responsibility (policymakers) that a problem truly exists." (p. 278) "A glance at state and national data since the second half of the 1960s clearly shows that the entire primary and secondary school system was -- and is -- in jeopardy." (p. 280) He advocates that the federal government collect achievement data and use this information as a gauge in making financial grants to the schools.

Bullock, R., and Stern, J., "Estimated SAT Summary Statistics for High School Cohorts, 1967-1974." Memorandum, September 3, 1974. Princeton, N.J.: Educational Testing Service.

To answer the question of whether the number of junior scores of students who took the test again as seniors was a source of bias in the SAT "testing-year" means, special tables were developed in which only the senior scores of persons taking the SAT as juniors and seniors, i.e., "known" repeaters, were retained. The effect of removing these junior scores from the means for 1966-67 and 1973-74 was to reduce the mean decline in SAT-V from 27 points to 21 points, and the decline in SAT-M from 17 to 12 points.

Conference Board of the Mathematical Sciences, National Advisory Committee on Mathematical Education, Overview and Analysis of School Mathematics, Grades K-12. Prepared with the support of the National Science Foundation. Washington, D.C.: Conference Board of the Mathematical Sciences, 1975.

This study of the way mathematics is taught in the American public schools concludes that recent declines in mathematics test scores are not caused by faulty teaching but are a result of other factors that have caused similar nationwide declines in reading, social studies, and science test scores. The report reasons that "new math" courses are blameless because they are fundamentally sound and also because the spirit and substance of such courses have not been extensively implemented in the classrooms. Instead, because of declines in achievement levels in other school subjects, it is asserted: "Scholastic performance is sensitive to broad school and societal pressures that are little influenced by choice of content or teaching method in special subject areas."

Farr, R., Tuinman, J., and Rowls, M., Reading Achievement in the United States: Then and Now. Report prepared for the United States Office of Education by Educational Testing Service and Indiana University, 1975.

The investigators were concerned with whether the level of reading competency had declined over the years and sought to discover whether a data base was available to answer the question. They found that publishers kept renorming data confidential, they could find no data on readability levels of children's books, and they could not obtain records on reading performance of Army inductees. Census reports, too, were a blind alley, since the definitions of literacy were totally inadequate. Thus the data base for the study was the following: (1) extant research literature on studies of changes in reading performance; (2) a survey of 100 public school systems, to locate records of reading achievement change over a number of years; (3) a survey of the 50 states, to locate statewide reading achievement data. Despite the extremely inadequate extant data bases, the authors are willing to draw the following (tentative) conclusions: reading scores, as reported in the literature, appear to be higher for groups tested now rather than "then." School system achievement data show gradual improvement in reading competency for high school students, from the beginning of the century up to 1965; there may, however, be a very slight decline in reading achievement after 1965.

Fasold, J., Hall, M., and Impara, J., Decline in Standardized Test Scores: A Widespread Phenomenon. Salem, Oregon: Oregon Department of Education, 1974.

A compilation of several articles on the theme of a general decline in the ability of high school students.

Flanagan, J. C., and Jung, S. M., Progress in Education: A Sample Survey (1960-1970). Palo Alto, California: American Institutes for Research, December 1971.

This study covered the reading comprehension of students in the Project TALENT sample. A 48 item reading comprehension test was administered in 1960 to more than 400,000 high school students representing a random sample of 1,353 different schools. Ten years later the identical test was administered to over 12,000 high school students in 134 schools, the sample being selected in the same way as the 1960 sample. Over this 10-year period, the mean score increased slightly from 30.81 to 31.25.

Foster, B. J., Statistics of Public Elementary and Secondary Day Schools, Fall 1972. Department of Health, Education, and Welfare Publication No. (OE) 73-11402. Washington, D.C.: United States Government Printing Office, 1973.

Harnischfeger, A., and Wiley, D., Achievement Test Score Decline: Do We Need to Worry? Chicago: CEMREL, 1975.

"Score, Test, and Pupil Evidence for Test Score Changes" (summary, pp. 69-72):

"For the past decade, nearly all reported test data show declines from grade five onward. The declines become more pronounced at higher grade levels. This pattern is obvious in all tested areas. ... A few test data first seemed to contradict these findings:

"The PSAT data of eleventh graders did not show a systematic decline. But we also found factors that explain the apparent nondecline. After correcting for scale drift, the PSAT data follow the general decline pattern from the mid-sixties to 1970 and from 1972 onward. The score increase in 1971 could be due to changes in the test-taking population, as in that year the PSAT was connected to the National Merit Scholarship Competition, resulting in a more selective test-taking group. On the whole, we conclude that the analysis of the apparently contradictory results shows that accounting for scale drift and change in tested student populations alters the PSAT score pattern to conform with the common mold.

"NAEP's finding of increasing Reading-Literacy performance for eleventh and twelfth graders (17-year-olds) contradicts all other test findings for this age group. Here we believe the key for explanation lies in the test contents and tasks. NAEP's items address basics far below the eleventh and twelfth grade curriculums, while the other tests are oriented toward more grade-level specific content. Typical grade-specific achievement tests and the NAEP assessment are not easy to relate. However, we should assert that the NAEP increase in Reading-Literacy performance is valid and important, even though it is not an achievement area for eleventh and twelfth graders and it may not ultimately contradict the findings of decline in more complex content.

"The changes in Science performance among secondary pupils are somewhat more puzzling... The college-bound group taking the ACT shows fluctuation in Science achievement over the period between 1965-1966 and 1973-74 of about 12 percent of a standard deviation even though there is no systematic trend. This fluctuation is smaller than the size of the comparable drops in English (27 percent) and Social Studies (34 percent), but is still sizable. The pattern for Science was a low period from 1966-67 to 1968-69 followed by a rise, a small drop, and a recent rise to previous highest levels. The same age group, in their performance on a very similar test, the ITED test of Natural Science, shows decline. Twelfth graders' performance, in the State of Iowa, showed a large decline between 1968 and 1970, with performance plateaus before and after. The decline, therefore, matches a period of rise in ACT performance. Twelfth graders on the NAEP also show large declines during an overlapping period (1969-73). The only known systematic discrepancy between these data sets is in the college-boundness of the ACT test takers. Perhaps science achievement has dropped precipitously among the noncollege bound, but not among those preparing for higher education.

"Another startling finding which we could not explain by means of characteristics of the tests or test takers is the large drop in verbal achievement scores for females.

"Although we have to conclude that test scores declined for all tested achievement areas for grades five through twelve over the past decade, we found no evidence for declines in lower grades (grades two, three, and four), but perhaps even slight increases in achievement scores. Our analyses of tests and test takers indicates that these findings are real, not artifacts....

"The School, a Search Arena for Causes of Achievement Declines" (summary, pp. 106-107)

"Features of children's school environments, analyzed as to their potential relevance as determiners of achievement test score declines, carry highly discrepant burdens of likely responsibility. Pupil mobility or desegregation seem not to have any explanatory power, as regional and state differences contradict the widespread character of the declines. Teaching staff characteristics, such as experience and education, are possible distal causes but existing data do not reveal any evidence that they play a relevant role in the current test score drops. Also, for some of the factors connected to school reorganization in the past decade, such as continuing high school consolidation and the crowded state of elementary schools following the postwar baby boom, we could not draw any straightforward conclusion of powerful determination of decreasing achievement test scores. This similarity arises from the simultaneity of their accompaniment by massive funding increases and the inconsistency of their likely impact with grade level differences in the decline. And we lack data showing changes in pupils' learning motivation.

"Evidence for other contextual factors seems to reveal potential contributory power in the explanation of test score decreases. One such factor is pupil absence rate, which has steadily increased over the past decade, resulting in smaller average amounts of schooling for pupils, but also burdening the teaching process considerably.

". . .But the strongest explanatory power seems to come from curricular changes. Our gross data indicate a considerable enrollment decline in

academic courses. Secondary pupils have been taking fewer courses in general English and mathematics. But also enrollment in typical college preparatory courses, such as algebra, first-year foreign languages, and physics, is decreasing...."

Hawes, G. R., "The Decline of the SATs." CHANGE, November 1972, pp. 17-19. Average test score levels in college entrance tests appear to be declining, yet many teachers feel that students are not less able academically than their predecessors. In a technical report for the Carnegie Commission on Higher Education, Paul Taubman and Terence Wales found that the average mental ability of college students, as reflected in IQ test scores, was higher in the early 1960s than was that of college students in the 1920s and 1930s, in spite of greatly increased enrollments since World War II. Yet in recent years the decline of SAT scores has been noted, a striking case being that of Horace Greeley High School in the well-to-do suburb of Chappaqua, New York. The school has kept records of the quartile distributions of four-year cumulative course-grade averages of graduating classes for the last 23 years. The median in each of the last six years has varied only from 82 to 84, fluctuating between 80 and 83 in the previous 17 years. Yet SAT medians have steadily declined from 1962's 575 to 1972's 522.

Hilton, T. L., Rhett, H., Broudy, I. L., Bower, C., Carter, M. M., Creech, F. R., and Echternacht, G. J., The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972. Final Report. Princeton, N.J.: Educational Testing Service, 1973.

Jencks, C., Inequality: A Reassessment of the Effect of Family and Schooling in America. New York: Basic Books, 1972.
Argues that school quality has little effect on achievement, drawing upon major surveys conducted since 1960: Census Bureau studies of social mobility and income distribution, Project TALENT's surveys of American high schools, the Equality of Educational Opportunity Survey, and many others.

Johnson, Charles E., Jr., "Changing Trends in College Enrollment." College and University, Fall 1973, pp. 35-41.
Documents the many changes in colleges in the recent past: in the college age population, enrollment rates, proportion attending two-year or four-year colleges, proportion attending public or private colleges, and college costs and financing.

Maeroff, G., "The Missing School Yardstick: Measuring Gains and Losses." New York Times, March 17, 1975.
A call for better systems of monitoring pupil progress. Also comments on the fact that science scores on the National Assessment of Educational Progress have "slipped" since the tests were first administered in 1969-70 (the readministration took place in 1972-73). The greatest decline was evidenced by 17 year old students.

Marco, G. L., and Stern, J., Investigation of the "Repeater" Hypothesis Concerning SAT Score Declines. Princeton, N.J.: Educational Testing Service, April 1974.
The decline in senior SAT-verbal means from 1967 to 1973 was studied after adjustments to senior means were made under three different assumptions

about the proportion of repeaters. The samples were three mutually exclusive groups: (1) April or May juniors who did not repeat the SAT in their senior year; (2) April or May juniors who repeated the SAT in November or December; (3) November and December seniors who did not take the SAT in April or May as juniors. The three assumptions were these: (1) all the nonrepeating juniors repeated the test in November or December of the senior year; (2) the percentage of April/May juniors who repeated the test in November and December was the same in 1967 -- i.e., 67.32 percent; (3) the percentage of seniors in November and December who were repeaters was the same as in 1968 -- i.e., 33.83 percent, the year in which the highest percentage of seniors were repeaters. Since not all the juniors would ever be expected to repeat the test in the senior year, the first assumption leads to an overestimate of the effect of repeaters, but the other two give more reasonable estimates. The results indicate the "repeater hypothesis" accounts for about three points of the 26-point SAT-verbal mean score decline of November and December seniors from 1967-73. The effect on the decline of all seniors or all juniors and seniors tested during an academic year would be even less.

McCandless, S. A., The SAT Score Decline and its Implications for College Admissions. Paper presented at the 1975 Western Regional Meeting of the College Entrance Examination Board, January 1975.

Several hypotheses that have been offered to explain the score decline are here considered in detail: (1) the possible invalidity of the test; (2) the decreased numbers of SAT "repeaters"; (3) the instability of the score scale; (4) changing test-taking patterns; (5) change in the college-going population (less academically able because of expansion of the candidate group); (6) regional change in SAT-taking and immediate college entry; and (7) progressive decline in developed reasoning ability among young Americans generally. None of the available data were found to be decisively supportive of any hypothesis. Concludes with a discussion of the implications for the college of planning to arrest or reverse a projected score decline for prospective applicants.

The article contains the following useful tables: (1) SAT Score Averages for 1972 and 1973 graduates; (2) SAT Score Averages, by Sex, for 1972, 1973, 1974 graduates; (3) SAT Score Averages by Grade Levels for the Testing Years 1965-66 through 1973-74; (4) SAT-Verbal Score Distributions, for Graduates of 1972 and 1973 Taking the Test; (5) SAT Score Averages for SAT-Taking Students from High School Graduating Classes, 1967-1974; (5b) for Males; (5c) for Females; (6) SAT Score Averages 1956-57 through 1973-74; (7a) Estimates of the Rate of High School Graduation, Immediate College Entry, and SAT-Taking among Both Sexes, 1959-74; (7b) for Males; (7c) for Females; (8) Estimated SAT Score Averages for All Immediate College Entrants, 1961 and 1968; (9a) Percentages of 12th-Grade Males Classified by SES, Ability, and 13th-Grade Enrollment 1960-61 and 1967-68; (9b) for Females; (10) Percentage Distributions of SAT-Takers by Region of Residence, 1959-68; (11) Percentage Distributions of Immediate College Entrants, by Region of Residence, 1963 and 1968; (12) Primary Families with Civilian Dependent Members, 18 to 24 Years Old and Percent Enrolled Full-time in College by Family Income: October 1967 to 1972; (13) Estimated SAT Score Averages for All Seniors, 1960-61 and 1967-68.

Modu, C. C., and Stern, J., The Stability of the SAT Score Scale. Research Bulletin RB-75-9. Princeton, N.J.: Educational Testing Service, March 1975.

The authors investigated the consistency of SAT scales from 1963-73. Two verbal and two mathematical sections that had been administered in the 1963 and 1966 administrations were readministered to a random sample of the 1973 SAT candidate group. The raw scores on 1963 and 1966 forms were placed on the reporting scale for a 1973 form. The raw-to-scaled score conversions derived in 1963 and 1966 for equivalent raw scores representing the same ability levels on these forms were compared with their corresponding scores on the 1973 reporting scale. The equating procedures used in the experimental equating were identical with those used in the operational equating of SAT scores during the period of the study. However, in the operational equating, several equatings intervened between the tests that were equated directly in this study. The results indicated that for both 1963 and 1966 equatings, the 1973 candidate group would have earned lower scores on both SAT-verbal and SAT-mathematical had the experimental equating results rather than the operational equating results been used in reporting the December 1973 scores. The average difference was 14 points for SAT-verbal and 17 points for SAT-mathematical scores. The study data do not provide an explanation for the discrepancy between the results of the experimental and operational equating over the period of the study.

Further research is suggested to determine the bias in different equating methods, the effect of various methods of reliability estimates on equating results, the effect of variations in the performance of succeeding generations of candidates on common equating sections, the effect of the choice of equating sections, and the effect of the choice of old forms to which each new form is equated.

Munday, Leo A., Declining Admissions Test Scores. Iowa City: The American College Testing Program, 1976.

Over the last 10 years the decline in test score averages has been experienced by both ACT and SAT, and the extent of the decline -- about 2 percent to 3 percent of a standard deviation per year -- has been similar for both programs. There is variation by subject matter field; the decline in ACT scores has been most marked in social studies, there has been no decline in natural science, and there has been some decline in English and mathematics. The decline in test score averages has been accompanied by an increased variability in the academic achievement of students tested. Declining admissions test scores are a national phenomenon, occurring in all regions and in nearly all states. The percentage of ACT-tested high scoring students has remained about the same over the last five years, while the percentage of low scoring students has increased. There are marked sex differences. Overall, the scores of men have not declined appreciably over the last five years, but the scores of women have declined considerably. This trend has been accompanied by an increase in the percentage of women among ACT-tested students.

Two possible explanations for declining admissions test scores were examined. The first concerned who takes the test and the second, how well prepared they are. To deal with these questions, it was necessary to use data from state testing programs and the National Assessment since ACT data describe college-bound students and not high school students generally. The results for two states showed that in recent years, different people have probably been taking the test, and that starting in the late 1960s and continuing to the present, high school students have likely been less well prepared than previously. In different years one factor likely has more influence than the other, and in other years they combine in their influence. If the limited data studied have applicability to the national scene, it is possible to conclude that in recent years the ACT score decline has been largely due to the changed pool of test-takers. For their part, colleges and universities would seem to have adapted to the declining admissions scores of their incoming students. (Author's summary)

"Newest National Norms Show -- Decline in Scores Continues." ACTivity, October 1975, 13 (4), 5.
The 1974-75 norms for ACT Assessment test scores show a continued decline, yet reported high school grades continue to rise. The 1974-75 mean composite score for the four tests that make up the test section of the ACT Assessment is 18.6 (based on a 1-36 scale), compared to 18.9 a year ago and 19.2 in 1972-73. The mean high school grade point average (based on a four-point scale for self-reported grades in subject areas covered in the ACT Assessment test section) is 2.91 in 1974-75, compared to 2.86 the previous year and 2.81 in 1972-73. Composite scores are higher for males (19.5 against 17.8 for females); females' grade point averages are higher than those of their male counterparts (3.00 for females, 2.81 for males). More junior are taking the tests. Of the 900,000 taking ACT Assessment during the past year, 23 percent were high school juniors. Last year they were 20 percent of the total and 18 percent in 1972-73.

Perry, D., and Swanson, E. O., "Decline in Minnesota College Aptitude Test Scores." In J. Fasold et al., eds., Decline in Standardized Test Scores: A Widespread Phenomenon. Salem, Oregon: Oregon Department of Education, 1974.
The Minnesota Statewide Testing Program annually offers a Scholastic Aptitude Test, measuring verbal comprehension and reasoning, to all Minnesota eleventh-grade students. In 1966-67, 98 percent of all juniors took the test, but in 1972-73, the figure dropped to 91 percent. Both the mean and median have been steadily declining since 1968-69. In that year the juniors had a mean of about 33.0; in 1972-73 the mean score was 31.0 on the same test form. The standard deviation was 13.3 in 1968-69 and 12.4 in 1972-73. (Minnesota mean scores for juniors tested in 1963-64 through 1965-66 and for juniors tested in 1966-67 through 1968-69 stayed essentially level.) The statistics are not sufficient to explain the decrease. Fewer students participate, but those who do not take the test are generally lower-ranking students whose participation would have further decreased scores. The decline is due to the decrease in frequency of high scores combined with increase of frequency of scores close to the average. Perhaps the most likely explanation of the decrease in average scores is that student attitudes and motivation have changed.

Peterson, I., "College Textbooks being Simplified to Meet Needs of the Poor Reader." New York Times, November 7, 1974.
Textbook publishers have recently lowered the reading levels of college textbooks. Most of the pressure for improved "readability" came from community colleges that have had hugely expanding populations. There

was, however, evidence of sufficient trouble with reading comprehension at four-year colleges to persuade the publishers to produce new readability formulas.

Schrader, W. B., *Test Data as Social Indicators*. Statistical Report SR-68-77. Princeton, N.J.: Educational Testing Service, September 1968.

This survey was undertaken to find out whether useful year-to-year comparisons of educational achievement in the U.S. could be made using existing test data. Tests were located that provided the necessary data for making the comparisons, and all the 186 specific comparisons of earlier with later performance were expressed in a common unit -- the percentile rank on the new norms of a student who scored at the 50th percentile rank on the old norms. Using the norms for the Iowa Tests of Basic Skills (grades 3-8), the Metropolitan Achievement Tests (grades 5-8), and the Stanford Achievement Test (grades 5-8), Schrader found slight gains for students in the lower grades, who would have been high school seniors in 1968, over same-aged groups, who would have been seniors in 1957, 1960, and 1963.

Schrader, W. B., and Hilton, T., *Educational Attainment of American High School Seniors in 1960, 1965, and 1972: Feasibility study*. Final Report. Project No. NIE-G-74-0050, 1975. Princeton, N.J.: Educational Testing Service, 1975.

Schrader and Hilton, after a study of the feasibility of equating the 1960 Project TALENT, the 1965 Educational Opportunities Survey (Coleman 1966) and the 1972 National Longitudinal Study of the Class of 1972 (Hilton et al. 1973), propose an analysis of mean change of these national samples over the period of the score decline on the SAT.

Schrader, W. B., and Jackson, R., *Preliminary Report on the 1974 PSAT/NMSQT Norms Study*. Memorandum dated March 24, 1975. Princeton, N.J.: Educational Testing Service.

Using data from PSAT/NMSQT administrations in 1960, 1966, and 1970, the authors found that PSAT mean verbal and mathematical scores had remained virtually constant over a 14-year period, even though there was a slight increase in the 1966 means. In 1960, the mean verbal score was 34.2, and in 1974, it was 34.3. In 1960, the mean mathematical score was 38.7, and in 1974 it was 38.6. The schools were randomly selected, but from schools who voluntarily cooperated. The sample was not completely random because all schools selected did not participate in the study. In 1960, only seven-tenths of the schools participated, and in 1974 only three-fifths of the selected schools participated. Accordingly, the findings may be questioned.

Start, K. B., "Thirty Years of Reading Standards in England." *Educational Researcher*, 1972, 1 (11), 8-9.

A 1970-71 national survey of England's 11- and 15-year-old children was undertaken to establish current reading standards and to relate them to similar surveys that had been conducted in 1938, 1948, 1952, 1956, 1961, and 1964. Tests were administered to a stratified random sample, with school type and size providing the strata and the within-cell sample fractions drawn by birthdays. Scores in 1948 were below the 1938 estimates but had risen steadily since. When the 1970-71 figures were added to those of the previous twenty years, the data indicated that reading standards of the children leaving English primary schools "were no better now than they were in 1960 and could well be less than the standards in 1964." An official committee of inquiry has been established to investigate how the teaching of English, training of teachers, and monitoring of educational attainment may be improved.

Stewart, E. E., The Stability of the SAT-Verbal Score Scale. Research Bulletin RB-66-37, CEEB RDR-66-7. Princeton, N.J.: Educational Testing Service, July 1966.
Raw scores on each of four old forms of the SAT-Verbal test were placed on the reporting scale for a newer form. The new data were collected as part of the December 1963 administration; data on the old forms were retrieved from files for the April 1944, January 1948, March 1953, and February 1957 SAT administrations. Results: (1) For both the 1953 and 1957 forms, the experimentally derived scales and the reporting scales differed by less than five points in the vicinity of the mean, thus good enough to conclude that the SAT-Verbal scale remained highly stable during 1953-63. (2) Although the standard errors of both scales are relatively large, it may be concluded that the score of a candidate tested in 1963 and scoring in the vicinity of 500 might be 20-35 points higher than the score of a candidate of equal ability tested in 1948. Since the scale as it existed in 1953 was kept through 1963, the shift in the meaning of the scale must have occurred between 1948 and 1953. Scores on SAT-Verbal reported in April 1944 may not have had the same meaning as those of December 1963, but the methodological limitations inherent in the experimental equating of those two forms preclude any reliable quantitative assessment of whatever differences may exist between the scale of the early 1940s and that of the early 1960s.

Tavris, Carol. "The End of the IQ Slump." Psychology Today, April 1976, pp. 69-74.
Tavris summarizes Zajonc's arguments (see below): Today's low test scores are owing to (preceded by) children born close together in large families. The "IQ Slump" will be over around 1980 because the score declines directly reflect the increased family size of the postwar years and the students' later birth orders. We are already beginning to see an upswing in achievement in lower grades.

Thorndike, R. L., "Mr. Binet's Test 70 Years Later." Educational Researcher, May 1975, 4 (5), pp. 3-7.
Traces the history of the test since its development, about 1905, in France, through its American revisions in 1916 and 1960, and renormings in 1937 and 1972, giving a qualified defense of the test's unitary, overall appraisal of level of mental functioning. In the 1972 renorming, the shifts in distribution of IQs were "dramatic," especially the increase of scores for preschoolers; moderate gains were also noted for adolescents. The high average values were impressive because the sample in 1972 included, for the first time, ethnic minorities. This rise is interpreted as evidence of the increase in verbal and visual stimulation of young persons, especially those of preschool age. It is noted that the largest improvement has taken place in preschoolers' performance on nonverbal, pictorial, perceptual, and memory items, rather than on those that involve a good deal of semantic content. Thorndike suggests that our conception and use of Binet's test be modified so as to accord with current conditions in the modern egalitarian, pluralistic world.

Wilks, S. S., *Scaling and Equating College Board Tests*. Princeton, N.J.: Educational Testing Service, 1961.

Zajonc, R. B., "Family Configuration and Intelligence." *Science,* 16 April 1976, pp. 227-236.

Variations in aggregate intelligence scores are closely associated with variations in patterns of family configuration, and these aggregate family factors are deeply implicated in the declining SAT scores as a special case of a general phenomenon that manifests itself also in a variety of national, ethnic, regional, racial, and sex differences in intellectual test performance.

Several findings are pertinent: (1) Intellectual performance increases with decreasing family size. (2) Children born early in the sibship perform better on intelligence tests than later children when intervals between successive births are relatively short. (3) Long intersibling spacing appears to cancel the negative effects of birth order and in extreme cases to reverse them. (4) In general, long intervals enhance intellectual growth. (5) The adverse effects of short intervals are reflected in the typically low IQs of children of multiple births. (6) In the special case of only children, the benefits of a small family are apparently counteracted by the lack of opportunities to serve as teachers to younger children. (7) Last children suffer that handicap too. (8) Absence of a parent is associated with lower intellectual performance by the children. (9) Temporal changes in family patterns such as birthrates, average orders of births, intervals between children, and family size are reflected in temporal changes in aggregate measures of intellectual performance. (10) Differences in family patterns between different countries, between different regions of the same country, and between ethnic or racial groups are also associated with differences in aggregate intellectual performance. (11) Males and females differ in average birth order, and this difference is reflected in aggregate intellectual performance scores.

The large decline in SAT scores is surely influenced by the influences listed above. The decline in SAT scores (over 1/3 S. D. in 12 years) cannot be a function of changes in family configuration alone because it is considerably larger than would be expected on the basis of a simple extrapolation from four national samples (The Netherlands, 1965; United States 1965 NMSQT scores; France, 1973 scores; Scotland, 1947 Stanford-Binet scores). It is not claimed that the confluence model generates a unique interpretation of all these facts. For each of them one could probably supply another reasonable explanation. The intellectual deficit of twins could have a biological basis, for example, and the higher intelligence of twins who lost their co-twins may involve unknown genetic factors. The drop in SAT scores may be due to a general decline in intellectual interests... . Future research will shed light on these questions. At the moment, however, the confluence model has the advantage of parsimony. And because it makes rather specific predictions, it can be readily verified... . (from the author's summary)

Distribution of SAT Scores to Colleges as an Indicator of Changes in the SAT Candidate Population

William B. Schrader
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

September 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

ABSTRACT

This study attempted to identify trends in the extent to which Scholastic Aptitude Test (SAT) candidates were applying to various groups of colleges in 1960-61, 1966-67, and 1973-74. Its relevance to the question of the SAT score decline results from the information it provides concerning the effect of two developments on the SAT candidate group: (1) the widening of access to higher education, and (2) the acceptance of either the SAT or the American College Testing Program test (ACT) for admissions by a number of colleges. The first of these factors might be expected to bring in new candidates of less than average aiblity, and the second might be expected to remove traditional candidates of more than average ability from the candidate group. Although the interpretation of the data is complicated, especially by the lack of knowledge about the relation between college choice and participation in the SAT, the results offer some reason to believe that both of these factors resulted in changes in test taking between 1966-67 and 1971-72 that could have contributed to the SAT score decline. The data do not permit a numerical estimate of the size of these possible effects.

INTRODUCTION

Nearly all students who take the SAT presumably do so because they are applying to one or more colleges that require, or accept, SAT scores as part of the admissions process. During the past 15 years many additional institutions have required the SAT, some (notably the City University of New York and Bowdoin) have dropped the requirement of SAT, and many others now accept either the SAT or ACT as fulfilling their test requirement. It is clear that these changes have had some effect on the composition of the group taking the SAT. This report describes an exploratory study concerned with shifts in the number of SAT score reports sent to various groups of colleges and universities in 1960-61, 1966-67, and 1973-74. Institutions were grouped according to type of institution as defined by the Carnegie Commission on Higher Education, pattern test requirements for 1973 applicants, state and region in which the institution is located, and public or private control of the institution.

This study was made possible by the fact that, for many years, the College Board has made an annual compilation of the number of SAT reports sent to each institution receiving a specified number of score reports. The study was initiated, despite a recognition of its limitations, because it was expected to throw some light on changes in the composition of the SAT candidate group during the period before the

---

Author's note: Henrietta Gallagher was responsible for the intricate task of preparing the data for this study, and Ingeborg Stiebritz was responsible for the data processing.

Student Data Questionnaire (SDQ) was introduced. In addition to the complications inherent in an effort to match data for colleges over years, two major limitations of this study seriously complicate the interpretation of the results: (1) Because candidates may and do apply to more than one institution, data on score reports are at best only roughly proportional to the number of candidates who sent reports to a designated group of colleges. (2) Data from this study have not been related to data about the college-going population from other sources.

DEFINITION OF INSTITUTIONAL SAMPLE

The basic sample of the study included all institutions in the 50 states and the District of Columbia that appeared in one or more of the following College Board Score Report Lists: 1960-61, 1966-67, or 1973-74. (For 1973-74, only those institutions which received 50 or more SAT or College Scholarship Service reports were considered. Data on score reports were obtained from the 1973-74 supplementary list only for institutions that appeared on the 1960-61 or 1966-67 lists.) Wherever possible, data were matched across years even though the name of the institution changed, and different components of the same institution were grouped consistently across years. During the period covered by this study, the multicampus university type of organization gained greatly in importance, a development that seriously complicated the data compilation effort.

GROUPING OF INSTITUTIONS

The preparation of useful statistical data on higher education is greatly hindered by the lack of an accepted national classification of institutions. The diversity and complexity of higher education and the marked changes in organization and in conceptions of postsecondary education during recent years have further complicated the problem of comparing statistical summaries. Fortunately, the Carnegie Commission on Higher Education (1973) has published a sophisticated classification of substantially all institutions of higher education as one part of its very broad program of studies in this field. This classification provided one very useful way of grouping institutions. In addition, it was thought that grouping institutions by test requirements would be meaningful. For this purpose, data from a guidebook published under the auspices of the New York Times (Mazel, 1972) were used. Finally, institutions were grouped by states and regions in order to detect possible regional differences in the use of SAT scores at the various times studied.

The classification plan developed by the Carnegie Commission on Higher Education is the basis for the main comparisons in this study. This plan assigned each institution of higher education on the comprehensive list maintained by the United States Office of Education to one of 36 categories. Each institution was assigned to one of 18 classifications on the basis of various characteristics, and the institutions in each classification were further classified as under public or private control, resulting in 36 categories in all. Of the 18 classifications, 4 included only institutions that award a substantial number of doctor's degrees, 4 included colleges and universities that do not award substantial numbers of doctorates, 1 included two-year colleges, and 9 included various kinds of specialized institutions.

The doctoral-granting institutions were classified into four groups, mainly on the basis of the amount of federal financial support for academic science that they received in 1968-69, 1969-70, and 1970-71 and on the number of doctorates that they awarded in 1969-70. Assignment to categories was based on decision rules, supplemented by judgment in some instances. In general, it can be said that both research universities I and research universities II awarded 50 or more doctorates in 1969-70, but research universities I received more federal support. All universities that were in the highest 50 in numbers of doctorates awarded in the period 1960-61 to 1969-70 were included in research universities I or research universities II. Doctoral-granting universities included universities that awarded 10 or more doctorates in 1969-70, but the doctoral-granting universities I group either awarded 40 or more doctorates in 1969-70 or received $3 million in federal financial support in 1969-70 or 1970-71 and awarded 20 or more doctorates in 1969-70. In all, 173 universities were included in the four classifications: 52 in research universities I, 40 in research universities II, 53 in doctoral-granting universities I, and 28 in doctoral-granting universities II.

Most four-year colleges and universities not included in the four doctoral-granting classifications were classified as comprehensive universities and colleges or as liberal arts colleges, mainly on the basis of the number of programs offered and on size. Public institutions with fewer than 1,000 students and private institutions with fewer than 1,500 students in 1970 were classified as liberal arts colleges. Some judgment was exercised in making the classification. The comprehensive colleges and universities were divided into two groups on the basis of number of programs offered and on size. Liberal arts colleges were also divided into two groups. Liberal arts colleges I included institutions that scored 5 or above on Astin's Selectivity Index (which offers a rough estimate of aptitude level) or were included in the 200 institutions that produced the most doctorates earned at 40 leading doctoral-granting institutions from 1920 to 1966 according to a study conducted by the National Academy of Sciences. In all, 321 institutions were classified as comprehensive universities and colleges I, 132 were classified as comprehensive universities and colleges II, 146 were classified as liberal arts colleges I, and 573 were classified as liberal arts colleges II. All but 28 of the 719 institutions in the two liberal arts classifications were private. All the 1,061 two-year colleges and institutes on the list were assigned to a single classification.

Of the 18 classifications, 9 were assigned to a variety of institutions that offered specialized programs only. These included theological seminaries, bible colleges, medical schools, other separate health professional schools, technological schools, business and management schools, art schools, music schools, design schools, law schools, and teachers colleges. One classification, designated other specialized institutions, was treated separately in this study because the service academies were included in it. The remaining 8 were grouped as "Other Institutions."

In presenting the results of this study, some grouping of the 36 categories was done in order to reduce the amount of detail in the tables.

With respect to the other ways of classifying institutions in this report, it need only be mentioned that four categories were used in describing test requirements in 1972-73, as follows: SAT required, SAT or ACT required, ACT required, and test requirements different for residents and nonresidents. The fourth category was used for convenience; it is possible that a more detailed analysis of this category would have yielded useful information.

RESULTS

Table 1 shows the number and percent of score reports in the three years studied that were received by various categories of institutions, as defined in the Carnegie Commission classification. For all institutions combined, the number of score reports almost doubled between 1960-61 and 1966-67, and then fell slightly between 1966-67 and 1973-74. Although the number of score reports received increased between 1960-61 and 1966-67 for all but one category, the amount of growth was especially large for comprehensive universities and colleges (385,000). From 1966-67 to 1973-74 only three categories (two-year colleges, public doctoral-granting universities, and other institutions) showed an increase. Four classifications (public research universities II, private comprehensive universities, liberal arts colleges I and liberal arts colleges II) showed decreases greater than 50,000. The change in composition of the population of score reports is shown more clearly in the percentages. The percentage of reports received by liberal arts colleges I dropped from 13.2 in 1960-61 to 7.8 in 1966-67 and to 5.6 in 1973-74. Because inclusion in this group depends in part on estimated selectivity, this decrease deserves special attention. Public comprehensive universities rose from 9.7 percent in 1960-61 to 18.4 in 1966-67 and 18.9 in 1973-74. Finally, it may be noted that the combined percentages for two-year colleges and for other institutions, including various specialized colleges, new institutions, and institutions such as hospital schools of nursing not traditionally included in lists of colleges, rose from 3.2 in 1960-61 to 7.0 in 1966-67 and 14.1 in 1973-74. If it is assumed that these institutions tend to attract relatively low-scoring students, this increase may have some relevance on the score decline.

The results shown in Table 1 indicate that appreciable changes have occurred in the distribution of score reports to various kinds of institutions. The extent to which these changes arise from shifts in student preference from one College Board college to another College Board college in a different classification cannot be determined from

Table 1. Number of SAT Scores Received by Colleges and Universities Grouped by Carnegie Commission Classification in 1960-61, 1966-67, and 1973-74

Classification	Control	Number of SAT reports (in thousands)			Percent of SAT reports		
		1960-61	1966-67	1973-74	1960-61	1966-67	1973-74
Research universities I	Public	189	347	345	13.0	12.1	13.0
	Private	152	192	171	10.5	6.7	6.4
Research universities II	Public	151	286	199	10.4	10.0	7.5
	Private	73	96	65	5.0	3.3	2.5
Doctoral-granting universities	Public	75	221	245	5.2	7.7	9.2
	Private	112	166	122	7.8	5.8	4.6
Comprehensive universities and colleges	Public	141	526	502	9.7	18.4	18.9
	Private	177	313	244	12.2	10.9	9.2
Liberal arts colleges I		191	224	150	13.2	7.8	5.6
Liberal arts colleges II		108	260	209	7.5	9.1	7.8
Other specialized (including service academies)		34	33	33	2.3	1.1	1.3
Two-year colleges		28	141	229	1.9	4.9	8.6
Other institutions (including unclassified and unknown)		19	59	145	1.3	2.1	5.5
Total		1,449	2,864	2,660	100.0	100.0	100.0

the data. However, the results do suggest that the SAT candidate group is being drawn from a broader group of institutions and that changing patterns of test requirements may be affecting the distribution of score reports. To the extent that the effects of these two factors can be discerned in the results, Table 1 may provide useful information about changes in the SAT candidate population.

Table 2 shows the relation between the Carnegie classification and test requirements for the 1,181 institutions having data on both classifications. It should be emphasized that only those institutions which appeared on the College Board score report tabulations are included in this summary. As a result, a number of institutions that

Table 2. Number of Institutions Requiring Designated Testing Patterns, Grouped by Carnegie Commission Classification

		Test Requirement[a]			
Classification	Control	SAT	SAT or ACT	ACT	Different for non-residents
Research universities I . . .	Public	15	5	6	2
	Private	18	3	0	0
Research universities II . .	Public	6	7	6	3
	Private	6	6	0	0
Doctoral-granting universities[b]	Public	17	11	16	4
	Private	19	7	3	0
Comprehensive universities and colleges	Public	182	64	104	10
	Private	63	62	11	0
Liberal arts colleges I		90	44	1	0
Liberal arts colleges II		154	213	53	2
Other specialized (including service academies)		1	3	0	0
Two-year colleges		4	7	2	0
Other institutions		24	22	5	0
Total		499	454	207	21

[a] Source: Mazel, E., *The New York Times 1973 Guide to College Selection*. New York: Quadrangle Books, 1972

[b] Because this table is limited to institutions receiving a substantial number of SAT score reports, the number of institutions requiring ACT is understated.

require ACT only may not be included. Fairly sharp differences in test-requirement patterns characterize the various Carnegie categories. It appears that the use of either the SAT or ACT for admissions was widely accepted by 1972, especially by private comprehensive colleges and universities and by the less selective liberal arts colleges. For the group as a whole, the number that required the SAT (499 institutions) was somewhat greater than the number requiring the SAT or ACT (454 institutions).

Table 3 shows trends in the number of SAT reports received by institutions classified by test requirement and by public or private control. It will be noted that there was a decline in SAT score reports from 1966-67 to 1973-74 both for private institutions that required the SAT in 1972-73 and for those that required either the SAT or ACT in 1972-73. For those requiring SAT, the decline was 20 percent; for those accepting SAT or ACT, the decline was 39 percent. For public institutions which required SAT, the decline was 1 percent, and for those requiring the SAT or ACT, the decline was 13 percent. Although it is obviously risky to make inferences about the SAT candidate population from these data, the results do suggest that an increase in the number of colleges accepting either the SAT or ACT may well have affected the percentage of high-ability students who take the SAT. For example, some students considering Grinnell, which accepts either the SAT or

Table 3. Number of SAT Score Reports Received by Institutions, Grouped by Testing Pattern and Control, in 1960-61, 1966-67, and 1963-64

Test requirement[a]	Year	Public institutions	Private institutions	Total
SAT	1973-74	726,292	619,499	1,345,791
	1966-67	735,777	778,842	1,514,619
	1960-61	294,079	534,363	828,442
SAT/ACT	1973-74	296,929	268,015	564,944
	1966-67	341,811	436,769	778,580
	1960-61	142,142	260,862	403,004
ACT	1973-74	95,603	16,102	111,705
	1966-67	114,538	12,690	127,228
	1960-61	45,376	5,259	50,635
Different for non-residents	1973-74	129,553	570	130,123
	1966-67	82,134	869	83,003
	1960-61	34,282	1,001	35,283
Total	1973-74	1,248,377	904,186	2,152,563
	1966-67	1,274,260	1,229,170	2,503,430
	1960-61	515,879	801,485	1,317,364

a. Source: Mazel, E., The New York Times 1973 Guide to College Selection. New York: Quadrangle Books, 1972

ACT, and the University of Iowa, which accepts ACT only, would probably take ACT rather than both tests.

Table 4 shows regional patterns in distribution of score reports. These data, unlike those of Tables 2 and 3, show results for all institutions included in the study. From 1966-67 to 1973-74, the number of score reports received by public institutions increased for all regions except the Midwest, but for private institutions the number declined in all regions except the Southwest. For all institutions, only two regions showed a shift between 1966-67 and 1973-74 greater than 10 percent; the Midwest showed a decline of 32 percent and the West showed an increase of 11 percent during this period.

Tables 5, 6, and 7 show detailed results for each state and the District of Columbia. These results are presented for the use of readers who may wish to examine the sources of the regional patterns in greater detail.

Table 4. Number of SAT Score Reports Received by Institutions, Grouped by Region, in 1960-61, 1966-67, and 1973-74

Region[a]	Year	Public institutions	Private institutions	All institutions[b]
Middle States	1973-74	454,157	372,150	863,882
	1966-67	431,160	486,703	923,257
	1960-61	171,877	331,030	504,097
Midwest	1973-74	233,715	134,441	370,467
	1966-67	299,989	250,490	546,181
	1960-61	108,058	145,027	249,978
New England	1973-74	153,556	236,683	407,925
	1966-67	153,328	240,304	397,327
	1960-61	63,281	171,403	235,336
South.	1973-74	326,271	164,892	507,333
	1966-67	310,471	217,908	533,412
	1960-61	130,725	101,752	233,713
Southwest . . .	1973-74	102,322	33,502	138,317
	1966-67	94,632	32,801	127,687
	1960-61	19,981	15,740	35,721
West	1973-74	280,348	86,996	372,130
	1966-67	231,885	102,828	335,736
	1960-61	115,940	74,233	190,454
All states . . .	1973-74	1,550,369	1,028,664	2,660,054
	1966-67	1,521,465	1,331,034	2,863,600
	1960-61	609,862	839,185	1,449,299

a. Regions correspond to areas for College Board Regional Offices
b. Includes institutions that were not classified as public or private

Table 5. Number of SAT Score Reports Received by Institutions, by Region and State, in 1960-61, 1966-67, and 1973-74: Public Institutions

Region	State	1960-61	1966-67	1973-74
Middle States	Delaware	3,348	7,179	10,883
	D.C.	0	170	954
	Maryland	17,108	28,542	46,583
	New Jersey	26,421	63,792	69,207
	New York	69,681	154,042	176,152
	Pennsylvania	55,319	177,435	150,378
Midwest	Illinois	12,239	30,301	23,086
	Indiana	15,348	77,624	59,059
	Iowa	3,796	4,540	3,910
	Kansas	7,269	5,133	3,709
	Michigan	23,723	82,564	51,903
	Minnesota	2,908	4,983	4,538
	Missouri	3,057	7,444	9,423
	Nebraska	747	2,891	14,742
	North Dakota	253	246	620
	Ohio	28,250	65,298	45,782
	South Dakota	297	241	508
	West Virginia	2,711	6,074	7,538
	Wisconsin	7,460	12,650	8,897
New England	Connecticut	16,419	37,723	37,110
	Maine	4,591	11,706	11,157
	Massachusetts	26,131	69,700	67,367
	New Hampshire	4,526	11,012	13,276
	Rhode Island	6,209	14,064	12,413
	Vermont	5,405	9,123	12,233
South	Alabama	5,120	9,508	10,005
	Florida	11,629	27,788	36,540
	Georgia	32,309	66,947	61,500
	Kentucky	3,155	12,210	7,509
	Louisiana	2,342	6,092	5,764
	Mississippi	1,250	2,408	2,320
	North Carolina	27,604	79,596	87,810
	South Carolina	13,174	30,784	29,110
	Tennessee	3,307	8,373	7,578
	Virginia	30,835	66,765	78,135
Southwest	Arkansas	713	3,098	2,344
	New Mexico	1,952	3,347	2,773
	Oklahoma	2,638	5,492	4,765
	Texas	14,678	82,695	92,440

(Continued)

Table 5. (Continued)

Region	State	1960-61	1966-67	1973-74
West	Alaska	213	309	924
	Arizona	5,119	9,682	10,723
	California	50,333	112,867	172,475
	Colorado	30,680	42,415	34,649
	Hawaii	1,067	10,006	14,469
	Idaho	708	3,203	2,278
	Montana	672	2,124	2,292
	Nevada	402	1,005	1,210
	Oregon	16,814	33,069	26,965
	Utah	1,297	2,405	2,455
	Washington	8,211	13,463	10,632
	Wyoming	424	1,337	1,276

Table 6. Number of SAT Score Reports Received by Institutions, by Region and State, in 1960-61, 1966-67, and 1973-74: Private Institutions

Region	State	1960-61	1966-67	1973-74
Middle States	Delaware	137	1,075	1,958
	D.C.	20,284	36,313	22,429
	Maryland	11,338	13,645	13,915
	New Jersey	28,884	47,418	66,086
	New York	158,992	221,406	145,073
	Pennsylvania	111,395	166,846	122,689
Midwest	Illinois	34,958	54,560	26,912
	Indiana	22,063	43,307	29,739
	Iowa	8,757	13,995	4,370
	Kansas	827	1,814	1,085
	Michigan	7,837	19,769	11,022
	Minnesota	8,234	17,622	6,840
	Missouri	7,723	14,963	9,221
	Nebraska	772	3,195	4,385
	North Dakota	0	0	0
	Ohio	38,744	57,572	30,169
	South Dakota	69	373	345
	West Virginia	2,159	5,980	2,734
	Wisconsin	12,884	17,340	7,619

(Continued)

Table 6. (Continued)

Region	State	1960-61	1966-67	1973-74
New England	Connecticut	19,387	35,164	62,046
	Maine	8,370	9,883	9,435
	Massachusetts	113,037	155,188	122,707
	New Hampshire	9,615	11,671	14,813
	Rhode Island	12,791	18,971	17,630
	Vermont	8,203	9,427	10,052
South	Alabama	1,495	6,674	3,395
	Florida	11,706	24,729	26,326
	Georgia	13,518	25,489	20,356
	Kentucky	2,143	4,921	3,651
	Louisiana	7,664	12,219	7,408
	Mississippi	219	779	467
	North Carolina	32,682	70,953	49,413
	South Carolina	8,970	22,149	19,332
	Tennessee	8,104	16,067	12,762
	Virginia	15,251	33,928	21,782
Southwest	Arkansas	545	1,407	878
	New Mexico	66	273	681
	Oklahoma	393	2,806	4,550
	Texas	14,736	28,315	27,393
West	Alaska	0	52	92
	Arizona	0	0	491
	California	49,604	64,291	61,941
	Colorado	7,875	11,618	7,037
	Hawaii	174	2,825	1,445
	Idaho	244	214	845
	Montana	105	168	326
	Nevada	0	0	70
	Oregon	8,236	10,775	7,186
	Utah	689	2,128	2,128
	Washington	7,306	10,757	5,435
	Wyoming	0	0	0

Table 7. Number of SAT Score Reports Received by Institutions, by Region and State, in 1960-61, 1966-67, and 1973-74: All Institutions

Region	State	1960-61	1966-67	1973-74
Middle States	Delaware	3,485	8,254	12,841
	D.C.	20,579	36,774	23,900
	Maryland	28,446	42,187	61,407
	New Jersey	55,355	111,357	142,804
	New York	229,251	379,422	335,162
	Pennsylvania	166,981	345,263	287,768
Midwest	Illinois	47,197	85,196	51,548
	Indiana	34,304	112,809	84,937
	Iowa	12,553	19,113	8,296
	Kansas	8,096	6,947	4,794
	Michigan	31,560	102,583	63,514
	Minnesota	11,142	22,791	11,482
	Missouri	10,780	22,407	18,958
	Nebraska	1,519	6,159	19,885
	North Dakota	253	246	620
	Ohio	66,994	124,249	78,351
	South Dakota	366	614	909
	West Virginia	4,870	12,054	10,339
	Wisconsin	20,344	31,013	16,834
New England	Connecticut	36,040	73,248	101,579
	Maine	12,961	21,639	22,707
	Massachusetts	139,586	228,013	201,554
	New Hampshire	14,141	22,745	29,104
	Rhode Island	19,000	33,132	30,635
	Vermont	13,608	18,550	22,346
South	Alabama	6,615	16,182	13,400
	Florida	23,335	52,746	66,510
	Georgia	46,743	94,529	85,446
	Kentucky	5,298	17,131	11,449
	Louisiana	10,006	18,311	13,172
	Mississippi	1,469	3,187	2,787
	North Carolina	60,286	151,516	139,438
	South Carolina	22,144	52,988	48,892
	Tennessee	11,411	24,516	20,657
	Virginia	46,406	102,306	105,582
Southwest	Arkansas	1,258	4,505	3,222
	New Mexico	2,018	3,620	3,454
	Oklahoma	3,031	8,298	9,315
	Texas	29,414	111,264	122,326

(Continued)

Table 7. (Continued)

Region	State	1960-61	1966-67	1973-74
West	Alaska	213	361	1,016
	Arizona	5,119	9,682	11,214
	California	100,167	177,625	236,781
	Colorado	38,555	54,033	41,960
	Hawaii	1,241	12,831	16,912
	Idaho	952	3,417	3,123
	Montana	777	2,292	2,618
	Nevada	402	1,005	1,280
	Oregon	25,101	44,349	34,699
	Utah	1,986	4,533	4,583
	Washington	15,517	24,271	16,668
	Wyoming	424	1,337	1,276

DISCUSSION

The fact that SAT candidate volumes have remained essentially stable during most of the period when scores were declining may be taken to indicate that progress toward widened participation in higher education has had little effect on the SAT candidate group. This view would seem to be reinforced by the similar experience of ACT during recent years.

It can be argued that the results of the present study lend credibility to an alternative hypothesis. In this view, the stability of candidate volume has come about because two major trends have tended to offset each other. The acceptance of either the SAT or ACT by many institutions has probably reduced the number of "traditional" candidates who take both tests, especially in the last five years. On the other hand, the broadened participation in higher education has probably brought additional students into both testing programs during this period. It should be emphasized that the increased access of minority students to higher education is only a small, though significant, part of this movement.

The alternate hypothesis might account for a substantial part of the score decline if the trends are fairly large and if the "nontraditional" students earn substantially lower scores on SAT, on the average, than "traditional" students. The data of this study do not permit a rigorous evaluation of this effect, because a shift in a student's college preference would have no effect on his or her membership in the SAT population if both colleges required SAT, for example, and because shifts in test requirements within a particular group of colleges may have tended to offset each other to some extent.

REFERENCES

Carnegie Commission on Higher Education, A Classification of Institutions of Higher Education. Berkeley, Calif.: The Carnegie Commission on Higher Education, 1973

Mazel, E., The New York Times 1973 Guide to College Selection. New York: Quadrangle Books, 1972

Television and the Test Scores

Wilbur Schramm
Distinguished Center Researcher
The East-West Center, Honolulu, Hawaii

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

August 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

INTRODUCTION

There is no persuasive evidence that children's and young people's use of television has caused the decline in Scholastic Aptitude Test (SAT) scores. Indeed, a social phenomenon as broad as this one is -- a steadily declining trend in college entrance examination scores as well as in other tests, some of which measure different cognitive skills and different grade levels -- over a 12-year period, would not likely be explained in terms of a single cause. If television has anything at all to do with the decline in test scores, it is probably not the cause, but rather a cause, perhaps part of a complex causal system.

This paper examines several questions that might throw light on the relationship:

1. Does the timing of television's growth in this country relate to the timing of the decline in the test scores in a way that suggests there could be a causal relationship?
2. Has the television behavior of young Americans been of a kind to suggest that it could have substantially affected the children's school work and presumably, therefore, their performance on academic tests?
3. Is there any hard evidence that television exposure has, in fact, affected academic performance?

The Timetable of Television

There were fewer than 100,000 television receivers in the United States at the end of the war in 1945; almost all of these were in the northeastern part of the country. In 1947, there were about 300,000; in 1950, about 10 million. Bogart (1956) has extracted these figures, describing the growth of television in the early years, from Market Research Corporation reports:

	September 1949	October 1950	July 1951	July 1952	July 1953	July 1954	July 1955
U.S. Total. . . .	6%	18%	27%	37%	49%	58%	68%
Region							
Northeast . . .	13%	35%	45%	59%	69%	77%	85%
South	1	4	13	17	30	37	54
North Central .	4	15	27	39	50	61	70
Mountain and Southwest. . .	*	3	10	16	31	40	52
Pacific	5	19	26	34	47	56	66
City Size							
Farm.	*	3%	7%	12%	23%	28%	43%
Under 2,500 . .	*	5	9	16	26	41	56
2,500 to 50,000	*	9	16	18	31	39	53
50,000 to 500,000. . . .	*	17	28	37	51	64	75
500,000 and over	14%	40	53	69	77	81	87

(continued)

	September 1949	October 1950	July 1951	July 1952	July 1953	July 1954	July 1955
Income							
Upper fourth. .	7%	24%	33%	45%	58%	70%	81%
Next fourth . .	7	19	32	41	55	63	76
Next fourth . .	6	18	28	40	50	59	69
Lowest fourth .	3	12	18	23	32	41	48
Education							
Grade school. .	4%	16%	23%	31%	43%	51%	62%
High school . .	7	22	34	45	57	65	75
College	6	17	24	38	48	61	73
Family Size							
1 and 2 members.	4%	12%	19%	28%	38%	48%	59%
3 members. . . .	6	18	30	41	52	65	73
4 and 5 members.	7	22	35	45	58	69	78
6 and more members	6	19	27	39	53	55	66
Age of Housewife							
Under 35. . . .	8%	23%	36%	44%	58%	65%	75%
35 through 44 years.	8	23	37	48	60	68	74
45 years and over	3	12	19	28	40	49	61
Presence of Children							
5 years and under.	7%	23%	36%	48%	54%	65%	75%
6 to 12 years .	7	22	35	46	56	65	76
13 to 20 years.	6	18	28	39	49	58	71
No children . .	4	13	21	29	37	50	61

* No data

This is an exponential curve of growth -- the familiar S-curve of diffusion of innovation. The television growth curve is distinguished from that of other technological innovations -- the telephone, automobile, and radio, for example only by its sharp slope. The growth curve extended throughout the 1950s, flattening out in the 1960s. About 94 per cent of all homes in the United States are believed to have installed television by 1965. By 1967, about 98 per cent of homes had their own sets. The percentages have not increased since then, although more receivers have been added as multiple sets in homes, portable sets, and installations for schools and other gathering places. Also, many black and white sets have been replaced by color. Thus, the curve was nearly flat until a year or two after the war of 1939-45, rose precipitously for the next 15 years, and has been essentially flat (because of saturation) since 1967.

A year of special interest to us is 1963, when the SAT scores first turned downward. In that year there were about 56 million television receivers in the United States, representing over 90 per cent of homes. About 3 million of these receivers were color. In other words, television was already close to saturating the American home market, and the changeover to color was well under way.

It apparently takes a period of time for television's effect, if there is one, to show up in children's school work. The first major studies (Himmelweit, Oppenheim, and Vince, 1958, in England; Schramm, Lyle, and Parker, 1961, in the United States and Canada) found no significant relationship between children's television exposure and their school performance when television was relatively new in the communities where they lived. NHK[1] researchers (NHK, 1969; Faru, 1971) found no significant relationship of this kind in Shizuoka in 1957 or 1959 when television was new there. But they went back to Shizuoka several times in the next ten years, and were able to observe a relationship in 1967 and 1969, when television was ten years old and almost all homes had receivers. Therefore, it is necessary to look at the timetable of television ten years or so before the decline in test scores became evident.

In 1953, for the first time, more than half the children in the United States were exposed to television in their own homes. Perhaps a quarter had been watching television at home since the age of three or four, and 80 to 90 percent had some exposure to television, if not in their own homes then in the homes of friends or relatives.

In 1963, therefore, more than half the young people who took the SAT had been exposed to television at home for ten years or more, 75 to 80 percent had been exposed for at least five years, almost all of them had television at home and watched it several hours a day.

What constitutes a critical mass of children, television, and time to make a (perhaps fairly small) effect reveal itself on an academic achievement test? At this time, there is no way to answer that question scientifically, but it seems reasonable to believe that such a critical mass would have been available by the early 1960s. A sufficient number of children and young people would have spent a sufficient number of hours watching television, there would have been sufficient substitution of television for other activities (reading, for example), sufficient practice in the cognitive and perceptual patterns of television, sufficient learning of television's content, to make a difference if there were going to be one. We can not claim, at least on the basis of this evidence, that the kind of effect we are talking about did indeed occur, but if it did the time would have been ripe for it by 1963.

Let us look at the nature of the television experience to try to decide whether it is of the kind to affect academic performance.

Television -- the Thief of Time

Probably no technological innovation in this century has so influenced time allocation in human societies as has television. A team of researchers sponsored by UNESCO in the mid-1960s, made 14 time studies in 12 countries -- two samples in the United States, two in West Germany, one in Latin America, and nine in East and West European nations. The results were reported by Szalai (1966), Robinson and Converse (1970), and Robinson (1971).

---

[1] Nippan Hoso Kyokai -- The Japanese Broadcasting Corporation.

From the resulting time-use diaries it was possible to see how the acquisition of different technological devices changed the time allocations of their owners. Surprisingly, owning an automobile cost the owner only about six additional minutes per day on transportation, although he obviously must have traveled farther. Owning one of the new "labor-saving" household appliances like clothes washers, dish washers, or vacuum cleaners made almost no difference in the amount of time devoted to house work. But when a family acquired television, the time its members spent on media was increased by an average of 58 minutes a day!

Americans in the sample spent more time on television than did people in any of the other countries studied, probably because more was available. However, when only the diaries of persons with television in their own homes were examined, the daily television time was almost the same across samples. And this despite the striking differences in programming and the amount of programming available!

In that study, television accounted for about 40 percent of the leisure time of Americans. The time they spent on television was about three-fourths the total time spent on mass media, indicating that large amounts of time were stolen from radio, movies, and perhaps newspapers and magazines. Among 37 principal activities on which data were recorded, television was third only to sleep and work as a consumer of adult time.

That television was displacing other media had been documented ten years earlier than the UNESCO study, when Coffin (1955) reported on the daily media behavior of respondents in Fort Wayne, Indiana. Television had recently come to that city, and individuals who had acquired sets were interviewed for time diaries some months before and again after its advent. The summary table was as follows:

	Before television was available in the home	After a receiver was purchased	Difference
Magazines.	17 minutes per day	10 minutes per day	-7 minutes
Newspapers	39	32	-7
Radio.	122	52	-70
Television	12	173	+161
Total.	190	267	+ 77

These television owners were transferring about 77 minutes per day of leisure time beyond what came from the media. Leisure time is not indefinitely expansible: those 77 minutes a day, approximately nine hours a week, must have come from some other activities. Where did they come from?

Some movies, of course; over the whole sample, perhaps an hour a week. But as the UNESCO figures show, the new time for television came also from sleep, from social time away from home, from reading books, from conversation, from travel and other leisure activities, and from religious activities.

Coffin's figures are for adults, but the same general pattern of displacement is also true for children.

Children's time with television has been documented now for more than 15 years, not only by the early large studies (Himmelweit et al., Schramm et al., previously mentioned), but also by a large body of research done under the auspices of the Surgeon General and reported in the fourth volume of the study's report (edited by Rubenstein, Comstock, and Murray, 1972). Therefore, we have a wide data base over groups and over time. And the figures are impressive.

By the age of two, most children in the United States have begun to spend some time with television. Before the age of six, most have become purposeful television viewers, with regular viewing times and favorite programs (Lyle and Hoffman, 1971; Murray, 1972; Gerbner, 1971). The modal viewing time, even in the preschool years, is more than two hours. Murray (in the report cited above), studying inner-city kindergarten and first grade children, found a range of viewing time from 5 to 42 hours a week, with a mean of 21 hours, 29 minutes. Stein and Friedrich (1971), studying nursery school children in Pennsylvania, reported mean weekly viewing times of 34.56 hours for boys, 32.44 hours for girls. Lyle and Hoffman (1971) found that a large sample of first grade children in the Los Angeles area averaged a little more than three hours a day in the first grade. Lest we mislead ourselves into thinking that this is peculiarly American behavior, it should be noted that Japanese children begin their television viewing at least as early and view at least as much as American children (NHK, 1969; Furu, 1971). And the same thing is probably true anywhere in the world where television is available as readily as it is in Japan and the United States.

During the first 20 years of life in America, the chart of television use is curvilinear. It is highest at the beginning of the teen years, and trails off after that. Lyle and Hoffman (cited above) found, for example, that their sample viewed slightly more than three hours an average day in the first grade, peaked at about four and one-half hours in the sixth grade, and averaged just under four hours a day in the tenth grade. Chaffee, McLeod, and Atkin (1970) reported that the senior high school students in their sample watched about one and one-quarter hours less, on the average, than their junior high school students. There is considerably more viewing by children beyond the first few grades on Saturday and Sunday than on school days. Lyle and Hoffman, for example, found that sixth graders viewed almost twice as long on Sunday as on an average week day. One more characteristic of television viewing is that, both for young people and adults, blacks view more than whites, blue collar families more than white collar families (for example, Greenberg and Dominick, 1969; McIntyre and Teevan, 1971).

And the amount of viewing among children and young people has been increasing. Average viewing times rose about an hour per day in the 1960s. Lyle and Hoffman, in 1970, repeated a part of the Schramm, Lyle, Parker study of 1959 (published 1961). These were the comparisions:

Median self-estimates of viewing time for a weekday

	1959 San Francisco		1970 Los Angeles	
	Boys	Girls	Boys	Girls
First grade	2:00	2:00	Not available	
Sixth grade	2:30	2:30	3:19	3:24
Tenth grade	2:45	2:48	3:03	3:13

Suburban sample

	1959	1970 Boys	Girls
Weekdays:			
First grade	2:18	3:32	3:26
Sixth grade	2:30	3:26	3:40
Tenth grade	Not available	3:13	3:26
Sundays:			
First grade	2:24	2:15	3:19
Sixth grade	3:54	6:37	6:38
Tenth grade	Not available	5:38	5:40

A corresponding increase in viewing during the 1960s has been reported in Japan. Comparing two surveys conducted by the NHK Radio and Television Culture Research Institute in 1959 and 1967, Furu (1971) noted that the total viewing hours for a four day period, including Saturday and Sunday, increased by an average of four hours for fifth grade boys, four and one-half hours for fifth grade girls, and five and one-half hours for seventh grade boys and girls.

The importance of this increase may be considerable. If the attraction of television for children was strong in the 1950s, it must have been no less in the 1960s and early 1970s. (Indeed it must have been no less in the 1960s and early 1970s.) In fact, it must have been growing during the years that the decline in college entrance scores began to be noticed.

These figures are probably somewhat overestimated. Bechtel, Achelpohl, and Akers (1971) installed videotape equipment in 20 homes to record time actually spent in viewing, as contrasted with diaries kept by viewers and with their estimates of the previous day's viewing. They found a great deal of time in which viewers' attention strayed from the television set, they engaged in other activity or went out of the room for a while; and the researchers suggested viewers' diaries probably overestimate actual viewing time by about 25 percent, and viewers' figures on the previous day's viewing overestimate it by 40 to 50 percent. A sample of 20 families is too small to be definitive, and the average difference may therefore be less than Achelpohl et al., have reported. But we must assume that such figures on viewing time as we have cited in the preceding paragraphs are somewhat larger than the true figures.

Suppose, however, that the figures given in those standard studies were as much as 25 percent overestimated. That would mean that children for 14 years or so before they take the SAT spend on an average between two and three hours, rather than between three and four hours, viewing television per day. In their first 16 years of life, therefore, and before they take a college entrance examination, they would have spent somewhere between 10,000 and 15,000 hours watching television. This is substantially more time than they spent in school, very likely more than they spent in play, in fact, more time than they spent doing anything else except sleeping.

Television and a Child's Life

What might have been the effect of this vast expenditure of time?

The first major American study of television and children was entitled Television in the Lives of Our Children, and ever since then (1961) we have been trying to understand what part the medium really plays in the lives of what sorts of children. That it is important to children, that it has an influence on their lives, we have no doubt; but we have been somewhat surprised at how much they take it for granted, how soon they become blasé about it. Television has become for today's children a common and expected part of life. If one of our children, 50 years ago, had spent one-sixth of all his time in an environment very different from his usual one -- for example, a country child in the city, a city child on the farm, a poor child in a millionaire's house -- we would have expected the experience to have had a significant effect on that child's fund of knowledge, his values, and his behavior. We should expect the same, now, of children who spend so much time in the very different, exciting, vivid world of television.

About a child's life with television we know at least three things with some confidence: it displaces other activities and in that way helps to create a different life style; a great deal is learned; and the experience develops a child's way of perceiving and learning that is different from the reading and discussion most classrooms emphasize.

From a large number of studies of children's television behavior, we know that children with television, as compared to those without it:

-- overwhelmingly think of television as their favorite medium, and as a result spend less time listening to radio or going to the movies -- functional equivalents for their television experience;

-- go to bed later;

-- are less likely to be read to by their mothers before they themselves learn to read;

-- spend less time reading when they do learn how to read;

-- spend less time playing with other children away from their own home.

In addition, television seems very close to them. They say (Lyle and Hoffman, 1971, table 64) they would be more likely to watch television than engage in other activities, for entertainment, for relaxation, and for relief from loneliness. The same study reported that television intrudes both into their play and their dreams. About half the children studied said that they sometimes "played out" what they saw on television, and 40 percent of them said they remembered dreaming about things they had seen.

Does television time interfere with homework? There is very little hard evidence. In trying to answer the question we are reduced mostly to children's own reports, parents' worries, and a few conjectures by teachers, most of these from the early years of television. Remmers' survey of teenagers (1953) found two out of five thought television interfered with their preparation for school work. Battin (1954) found that a fifth of the children she studied said that television interfered with their homework. Lewis (1951) found that 69 percent of children in television homes said they did their homework before they watched TV. Witty (1950 et seq.) in his Evanston surveys found many teachers and parents critical of television's suspected effect on study and homework. NHK's studies of attitudes toward TV (see NHK, 1969) found that 55 percent of children in the fourth grade, 58 percent in the seventh grade, and 71 percent in the tenth grade said they thought television disturbed their homework. As contrasted with that, only 4 percent in the fourth, 4 percent in the seventh, and 5 percent in the tenth grade, felt that television was a useful source of knowledge. On the other hand, 46 percent of children in the fourth grade, 36 percent in the seventh, and 35 percent in the tenth grade felt that the printed media were sources of useful knowledge. One other possible effect of media on homework, little studied as yet, is the apparently growing habit of doing homework with radio turned on to music or television turned on to an entertainment program. It has been estimated that about half the children who are assigned homework study with the radio for background, and an unknown but increasing proportion use television for the same purpose.

Television, therefore, may be affecting the amount and quality of homework. If so, along with the later bedtimes, it might make for some sleepy children in school the next morning; and this, together with the lesser amount of reading or being read to, might affect school performance.

But the greatest direct effect of television time on children may well be what it teaches them to do, rather than what it keeps them from doing.

Nothing is more characteristic of studies of children and television than reports by the often amazed authors of how much children learn from the experience. McLuhan (1964) and others writing in this vein have sometimes asked why educators concern themselves with school television when the real television school goes on all the time, outside school, with the teachers paid chiefly by NBC, CBS, and ABC.

"Television can be considered to be a window on the world, a school if you will, through which the child perceives his society and then learns from repeated example to cope with the vicissitudes of living," concluded one of the research coordinators of the Surgeon General's study (Murray, 1972). This sums up the general impression of many scholars. Yet it is a peculiar kind of world that a child sees through the window, and a peculiar kind of school that teaches the child to cope. It is a world of instant heroes and of instant news events that are often forgotten before one learns how the story "comes out." It is a world full of violence, and yet no one seems to feel any pain or spill any blood. It is a world in which most of the people are stereotyped and one dimensional. Taxi drivers all have the same set of characteristics; women are locked into a set of occupations and behaviors that are more characteristic of women a half century ago than women today; cops are all of one kind, and so are crooks, and the crooks usually have quite a good time until they are caught, which is usually just at the end of the broadcast. Power and prestige seem to belong to those who command violence, those who are beautiful, those who can be entertaining and handsome before a camera and speak well into a microphone. Ideas and policies are not something to be explored and debated and studied in depth, but rather something to be dashed off with a summary sentence or an example. News is something that goes with a good picture. And for that matter, children and young people do not view much news anyway. McIntyre and Teevan (1971) reported that less than 20 percent of their teenage sample were news viewers, and Lyle and Hoffman (1971) found that their sample -- first, sixth, and tenth grades -- almost totally avoided news programs in favor of more exciting entertainment.

It is a curiously limited world to which most children are introduced, most of the time. Michael Arlen in McCall's (February 1971, quoted in Lesser, 1973) said what he thought of it without mincing words: "To pass through childhood," he wrote, "when one's imagination, one's sense of the possibilities of the world is developing, and have it touched only, or mostly, by a piece of boffo-funny mechanistic claptrap such as McHale's Navy seems a damn shame."

In general, it is a world that would seem not to contribute greatly to intellectual growth. A bemusing section in the Schramm, Lyle, Parker book (1961) compared learning by children in two Canadian towns, one with television, the other without it. The children in "Teletown" came to the first grade with a vocabulary, measured on the Lorge tests, about a year in advance of those from "Radiotown" (the one without television). In the early school years the children who had experienced television were more familiar with names of national and foreign leaders, with world capitols, with places like "Broadway and 42nd Street" or "the White House." By fourth grade, however, the vocabulary advantage no longer existed. By sixth grade, students from the town without television demonstrated that they had less information about public affairs, although more about entertainment personalities.

Thus, although a great deal of incidental learning does result from television, it is not typically the kind that raises scores on achievement tests. Behaviors, customs, clothes are copied, names are learned, phrases are added to one's vocabulary, skills are learned by modeling (how to catch a baseball or block for a quarterback, how to make love, how to use a hand gun). But the experience usually makes little contribution to reading or mathematics skills. The pupil has no opportunity to ask any questions, or talk things over with his teacher. The homework he is assigned (by implication) is to practice what he finds attractive in what he has seen, and to buy or persuade his parents to buy items that are advertised.

It is necessary, therefore, to think in two different ways of the time a child spends on television. It keeps him from doing certain things he might otherwise do that would contribute to his performance in school. It probably makes his bedtime later, causes him to read less and be read to less, and in some cases interferes with study at home.

It leads him to spend a great deal of his leisure time in an absorbing world which, for most children, is intellectually limited and is not the kind of preparation for doing well on an SAT test. We have no evidence that permits us to say that these experiences do affect test performance. We can only say that they might.

Still another aspect of children's use of television may be related to academic performance, although we know very little about how it works. This is the shifting of a major part of a child's learning time from the linear digital encoding associated with reading to the iconic simultaneous kind of encoding which is a chief part of watching television. McLuhan (1962, 1964) and others have written speculatively about the possible effect of this change upon personality and society; and some preliminary work has been done on the different encoding functions of the right and left hemispheres of the brain. This is not the place to go into that. Up to this point in research we can say very little with assurance about what may be happening, beyond the fact that children are now getting 15 to 20 hours a week more practice in learning from pictures than they used to, and a little less practice in learning from print.

But the fact remains that after 300 years of exercising the left hemisphere of the brain (the logical, verbal, linear side) our young people have now reversed the balance and are emphasizing the right hemisphere (the visual, affective, simultaneous one). This may be the most important of all the effects we have mentioned. Some signs are beginning to appear. A five-year evaluative study of El Salvador television (for a summary, see Mayo, McAnany, Hornik, 1976) found few significant differences in cognitive learning, on curriculum-referenced tests, between television and control groups, but did find a pronounced and increasing difference in the performance of the two groups on general ability, measured over five years. The major contribution to the increase in general ability scores came from the nonverbal parts of the test. Similarly, the National Assessment tests have revealed a changing style in writing that derives less from literary models and print than from the speaking styles common to the broadcast media. These are only straws in the wind. They do not prove that television has had anything to do with the decline in test scores. But this possible effect of television on learning is something we must watch and study during the years to come.

Cause and Effect

We have worked, so far, in the realm of the perhaps and the possible. Was the timing such that there could have been an effect of television on the declining test scores? Was children's television behavior a type that might have affected their behavior in the classroom, and subsequently on national tests? Now we must ask what hard evidence exists that there actually has been such a relationship between television behavior and academic behavior.

The problems of designing research on that relationship include two that are of considerable practical importance. It is almost impossible now, in this country, to find a suitable control group without television. In fact, it was already impossible in 1958 when this author was scouting for comparable cities or towns of at least 5,000 people, some of whom would not yet have television so that children who had television could be compared with those who did not. One town was found that received no broadcasts directly, and had only recently installed cable for which there were as yet only a few hundred subscribers, all of them readily identifiable from the cable company's records. But a little scouting around the town revealed that a very large proportion of the children seemed to watch television regularly. Further investigation revealed: (a) a number of potential subscribers had connected surreptitiously to the cable, (b) others had bought receivers and pointed their aerials at a flat mountain from which they could receive fuzzy reflections of a broadcast that originated more than 100 miles away, and (c) a high proportion of the children spent several hours a day with friends or relatives who did have television. After a few experiences of this type, my colleagues and I went to two remote towns in western Canada that represented the best experimental-control combination we could find.

That kind of experimental comparison was possible in England in 1956 when Himmelweit and her colleagues were studying the effect of television, and (just barely) in British Columbia in 1958 when data for Television in the Lives of Our Children were being gathered. It was still possible in the late 1950s in Japan when the first Shizuoka studies were being conducted. It is no longer possible, nor has it been for some years, in any of those countries. When Furu made his major study in and near Tokyo at the end of the 1960s he had no such control group, and was restricted to comparing heavy viewers with light viewers. This would have been the only comparison available in this country to the scholars who worked on the Surgeon General's study in 1969 to 1971. And comparing the effect of watching a lot of TV with that of less TV is obviously not the same thing as comparing the effect of TV with that of no TV.

The studies for which a control group without TV was available (Himmelweit, Oppenheim, and Vince, in England; Schramm, Lyle, and Parker, in Canada; and the early NHK studies in Shizuoka, Japan) were done when television was relatively new in the experimental areas, and, as mentioned previously there is good reason to think that it takes some time for the impact of television on school performance to become evident. This puts the researcher in a cruel bind: apparently he cannot demonstrate an effect when a suitable control group is available, and when the effect should be visible there is no control group of the kind he needs!

A second problem with research on this topic: it is fairly clear that the relationship of television viewing to school performance is not a simple one. A simple correlation is probably not sufficient. For example, Shuttleworth and May (1933) in their Payne Fund study of children and films found that heavy movie goers averaged lower than others in both school grades and deportment. But readers asked quite properly what this meant. Was it that heavy movie going caused children to study less or be sleepy in the mornings, or for some other reason to do less well in school? Or that frustration in school drove some children to try to forget their dissatisfactions in the movie theater? Or that some third variable or variables (intelligence, perhaps) led less gifted children both to do less well in school and be more willing to escape into movies? This is the problem of interpretation that television researchers face today when they find a negative correlation between amount of television viewing and school or test grades.

For these and other reasons the early studies of children and television did not report very striking effects of viewing on school work. Witty in his series of Evanston studies (fifth report, 1954) found that the bottom fourth of children in grades 3 to 6 (in respect to academic records) averaged 26 hours a week viewing television, whereas the top fourth averaged only 20 hours. A study by Xavier University in Cincinnati in 1950 (Xavier, 1951) matched children in TV and nonTV homes on their tested mental age, then compared their performance on the Metropolitan Reading Test and an Arithmetic Proficiency Test. There were no significant differences between the scores of children in TV and nonTV homes, nor between TV homes in which parents strongly controlled their children's viewing and those in which viewing was only lightly controlled.

Himmelweit's study in England came at a time when control groups were still possible. She gave teachers a form on which they were asked to compare the school performance of 172 pairs of children for several subjects, without being told which child was a television viewer and which one was not. The five-point scale she used produced these judgments by the teachers:

Control child much better than viewer	.24%
Control child a little better than viewer	.23%
Control child and viewer much the same	.17%
Viewer a little better than control child	.23%
Viewer much better than control child	.13%

Thus, 47 percent of the control (nonTV) children were judged as doing better than the controls; 36 per cent of the controls were doing better than the TV group. This trend toward better school work on the part of the controls was not statistically significant.

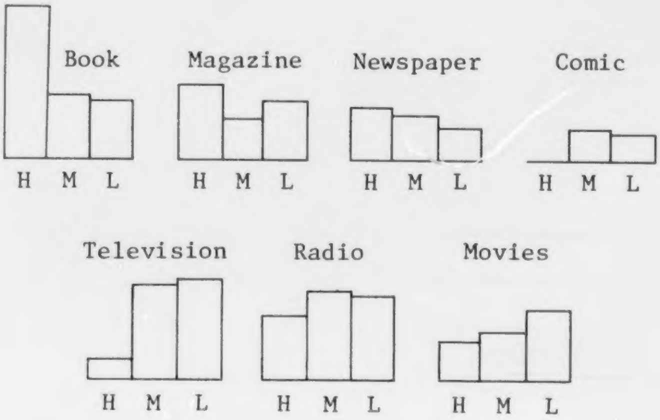
Himmelweit's summary:

> "On the whole we find that viewers are more or less able to hold their own, generally scoring just a little below their controls in the same classroom. This seems to be true for all sub-groups of age, sex, class, and intelligence, except for the brighter children. In both age groups these seem to fall behind a little more often.
>
> "Television had not led viewers to take a special interest in any one school subject; nor were they markedly better or worse at one subject rather than another. The very frequent showing of information programmes about nature, other countries, and current affairs had made little impact here; the viewers' performance at school in these subjects was no better than that of the controls. Television seems to be of no particular help to the child in making the most of his intelligence: it causes neither over- nor under-achievement.
>
> "As a social problem the effect of television on homework is slight." (pp. 310-311, 1958).

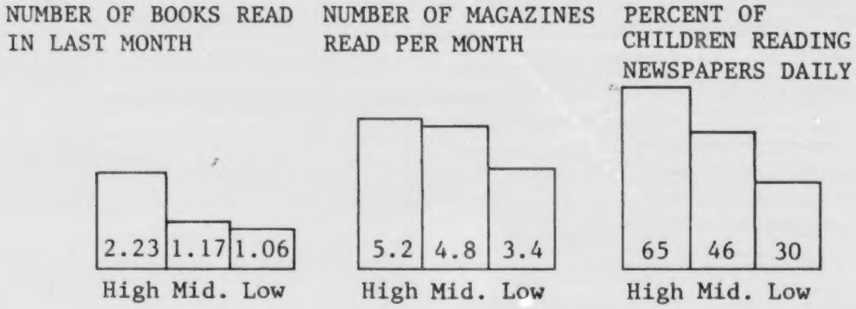
It must be remembered that this study was done when television had only recently begun to come into the communities studies. One of the more interesting findings is the suggestion of a greater effect upon the brighter children. This interaction of effect and intelligence came up again in the Schramm, Lyle, Parker study (1961), and was front and center in Furu's major Japanese study (1971).

The American study had an experimental and a control group in Canada, and made several surveys of communities where television was already in general use, in the United States. The experiment in Canada found no significant difference either in sixth or tenth grade school performance as between the town with and the town without television. It did find some differences in incidental learning, to which we have made reference earlier in this paper. That is, children with television came into the first grade with larger vocabularies -- both general vocabulary as measured by a standardized test, and special vocabulary as measured by a test of names and events in the news and in general discussion. They came into school also with a somewhat wider spread in general knowledge, but this seemed to be in the form of details, names, specifics, not general concepts or understanding or process and relationship. Heavy viewers in the first grade showed more learning of the specified kinds than did light viewers. However, all the advantage seemed to be lost during the first few years of school. At about the age of 12 there came a turning point in children's use of television and in its probable effects.

At this so called turning point, the importance of intelligence in the equation becomes readily apparent. The brighter children tend to be the heavier viewers during the first eight or ten years of life, but about age ten or from thereafter they turn to heavier use of print, and typically become the lighter viewers of television. The heavier viewers among teenage children tend to be those best able to identify singers and band leaders, the lighter viewers those best able to identify writers and statesmen. In other words, the fast start in general knowledge and vocabulary that television apparently encourages fails to last very long, and at about age 10 or 12 the situation reverses and the brighter children, those more interested in serious learning, begin to make greater use of print. Furthermore, television's prestige falls considerably. This can be illustrated by two charts from the Schramm, Lyle, Parker book (pp. 80 and 83), representing a large sample of eighth-grade children in several parts of the united States:



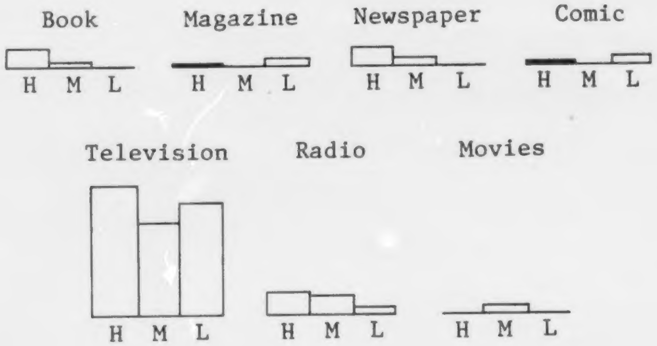
Prestige of media with eighth-grade children of different mental abilities (high, middle, low).



Reading of printed media by eighth-grade children of different mental abilities (high, middle, low).

The book we are discussing does not claim causation. It finds no significant relationship in the experimental situation. In the samples where control groups were infeasible, it discovers relationships between heavy viewing and family conflict, social norms (for example, a negative relation with the value placed on deferred gratification), personal adjustment and peer group relations, and, of course, intelligence. The relationship of televison to school performance is therefore part of a rather complex interactive relationship, not a simple matter of cause and effect. But it was clear even at the beginning of the 1960s that intelligence (and consequently in most cases higher performance in school) was associated, after age 10 or 12, with less use of television, heavier use of print, less prestige for television, and, in fact, some sense of guilt for spending a large amount of time with television. Television was regarded by older children as a source of entertainment and relaxation, and not much valued, for the serious public affairs and cultural programs it carried.

Yet, lest we make too much of this loss of prestige, it is necessary to cite one more finding from the 1961 book. Asked which medium they would miss most if they had to give it up, eighth graders gave these answers:



Percentages of eighth-grade children who would most miss each of the media, by mental ability (high, middle, low).

Thus, regardless of the fact that they thought of television as mostly useful for entertainment, the brighter children, even more than the others, said they would miss television most.

Furu's study, conducted in 1970 in Tokyo and its environs, made use of many of the scales and concepts of the Schramm and Himmelweit studies, and was able to go farther than the others in analyzing the elusive relationship between amount of viewing and school performance. Using a cluster sample with the school as unit, he selected, by systematic random sampling, 1800 tenth graders, 560 seventh graders, and 510 fourth graders. He secured measures, not only of viewing, print media use, and leisure time use, but also intelligence, creativity (a special Japanese test), adaptability (mostly questions on personal adjustment), social norm (immediate or deferred rewards), positivity (positive vs escapist behavior), parent-child conflict, reference group (parents' or peers' social norms), group membership (close friends or family), viewing situation, self-control of TV viewing, and the like.

This was a complex study and we cannot discuss it at length here. The point of greatest interest to us at the present time: Furu found, on his analysis of variance, a significant relationship between amount of viewing of television and school achievement scores in each of the three grades he studied. Part of his summary was as follows (pp. 150-52):

1. For each grade the school achievement score for high-intelligence groups was naturally higher than for the low-intelligence groups . . .
2. The school achievement score of the high-TV group was lower than that of the low-TV group for every grade . . .
3. The effects of print exposure were significant only for the seventh grade . . .
4. The effect of the interaction of intelligence and TV was not significant in any grade. This means that the effect of heavy viewing on school work was not affected by intelligence.
5. Regardless of the level of intelligence, the record of the high-print group was higher than that of the low-print group. In the tenth grade the difference was not significant . . .
6. In each grade the record of the print-type was higher than that of the TV-type . . .

The nature of his analyses of variance can be illustrated by the tenth grade analysis, which comes closest to the age when our own students take the SAT:

Source	df	MS	F	P
Intelligence.	1	112.541	9.730	.01
Television.	1	53.245	4.603	.05
Print	1	13.411	1.159	
Intelligence times TV	1	19.941	1.724	
Intelligence times print.	1	0.590	0.051	
Intelligence times TV times print	1	1.153	0.100	
Error	485	11.566		

n = 493

These findings seem to establish a close and consistent relationship between higher-than-average viewing of television and lower-than-average school achievement. But Furu is too careful a researcher to stop there. He assumed that a number of other variables are also important in determining the quality of school. Therefore, he introduced some of his other measures -- intelligence, adaptability, creativity, positivity, parent-child conflict -- along with a measure of TV and print behavior first used by Schramm, Lyle, and Parker, into a Hayashi multi-way analysis of fourth and seventh graders, using school achievement scores as dependent variables. This type of analysis combines multiple and partial correlations, and aims to determine whether one of the variables is significantly correlated with the dependent variable regardless of the effect of the other variables. The measure of TV and print behavior creates four groups by dichotomizing the subjects in different combinations above and below a midpoint. Thus:

High users -- both TV and print above midpoint
TV-type -- TV above midpoint, print below
Print-type -- print above midpoint, TV below
Low users -- both TV and print below midpoint

His purpose was to find out whether the TV-print difference was significantly related to school performance when the effect of the other variables was held steady. This is not quite the same as examining only the effect of high viewing of TV, but in a sense represents a purer measure.

By chi-square, all these variables are related to school work at a highly significant level (TV-print behavior, for example, at .001 and .005 in the two grades respectively). But when the partial correlations are obtained, the results are somewhat different. Intelligence, of course, is the chief correlate, as may be judged from the analysis of variance. But TV-print is not significant at the .01 level in either the fourth or the seventh grade (it apparently is significant at .05 in the fourth grade). Creativity, positivity, and parent-child conflict are significant at .01, along with intelligence, in the fourth grade; social class, adaptability, creativity, and social norms, along with intelligence, in the seventh grade. He says: "We can conclude that the lower achievement of heavy viewers is due not to the direct effects of TV but to the intervening effects of such predispositions as intelligence, creativity, adaptability, and so forth" (p. 163).

In Conclusion

Where does this review leave us? A bit frustrated and puzzled, perhaps, because we do not have clearer and more decisive evidence. Yet there are some conclusions that seem reasonable:

1. Television viewing is related to school performance and therefore very possibly to test scores: higher-than-average amounts of viewing, other things being equal, tend to be associated with lower-than-average academic performance.
2. When television comes into a family, it tends to reduce the amount of reading, and therefore provides less practice in that skill and less use of the cognitive and cultural material that print provides.
3. Television takes an enormous part of a child's life, and for most

children above the ages of nine or ten this activity is not very stimulating intellectually. Viewing television often gives a child a fast start with vocabulary and general knowledge, but in the early years of school the child outruns the intellectual level of television and uses the medium mostly for entertainment and relaxation.

4. A number of studies have found that television interferes somewhat with homework and keeps a number of children up late.

5. Television has shifted the balance of a child's practice away from learning from print (linear, verbal) to learning from pictures (simultaneous iconic). Most achievement tests, of course, depend largely on the former.

6. If television has anything to do with the kind of academic performance represented by the SAT scores, the early 1960s would be about the time we should expect the effect to show up. Actually the decline began in 1963.

7. However, a relationship is not necessarily a direct effect. The relationship to school work and probably to test scores is highly likely; the effect is neither proved nor disproved. However, considering the large number of powerful personal and social variables involved in academic performance -- intelligence, personal adjustment, relations with parents and peers, social norms, home and school situation, and so on -- would it not be extremely unlikely if we were to find a strong effect of television viewing independent of all these other variables? Furu calls them intervening variables, meaning that television acts through the others upon academic performance. It might be better to call them interacting variables, meaning that they act and react with and upon each other. This is to say, if adjustment affects television behavior, televison behavior may also affect adjustment, and both affect and are affected by social relationships; and just as they have something to do with performance in school, so the school experience may affect adjustment and social relations and the amount and kind of television one watches.

We can say that television is one of a group of strong variables that interact with each other and with school performance. On the degree to which television contributes to that system of interactions it is best to suspend judgment. But it is safe to say that whatever television viewing contributes, on the average, to a high school student's academic record or his performance in college entrance examinations it is more likely to be on the downward than the upward side.

BIBLIOGRAPHY

Battin, T.C., Television and Youth. Washington, D.C.: National Association of Radio-Television Broadcasters, 1954.

Bechtel, R.R., Achelpohl, C., and Akers, R., "Correlates Between Observed Behavior and Questionnaire Responses on Television Viewing," in Rubenstein, E.A., Comstock, G.A., and Murray, J.P., eds., Television and Social Behavior. 4 vols., Rockville, Md.: National Institute of Mental Health, 1972.

Bogart, Leo, The Age of Television. New York: Frederick Ungar Publishing Company, 1956.

Chaffee, S.H., McLeod, J.M., and Atkin, C.K., "Parent-Adolescent Similarities in Television Use." Paper presented at meeting of Association for Education in Journalism, August 1970, Washington, D.C.

Coffin, T., "Television's Impact on Society," American Psychologist. 1955, 10, pp. 630-41.

Furu, Takeo, The Function of Television for Children and Adolescents. Tokyo: Sophia University, 1971.

Greenberg, B.S., and Dominick, J.R., Television Behavior Among Disadvantaged Children. East Lansing, Michigan: Michigan State University Department of Communication, 1969.

Himmelweit, H., Oppenheim, A., and Vince, P., Television and the Child. London: Oxford University Press, 1958.

Lesser, G.S., Children and Television: Lessons from Sesame Street. New York: Random House Inc., 1973.

Liebert, R.M., Neale, J.M., and Davidson, E.S., The Early Window: Effects of Television on Children and Youth. New York: Elmsford Press, 1973.

Lyle, J., and Hoffman, H.R., "Children's Use of Television and Other Media," in Rubenstein, E.A., Comstock, G.A., and Murray, J.P., eds., Television and Social Behavior. 4 vols., Rockville, Md.: National Institute of Mental Health, 1972.

Mayo, J., McAnany, E., and Hornik, R., Educational Reform with Television: The El Salvador Experience. Stanford, California: Stanford University Press, 1976.

McIntyre, J., and Teevan, J., "Television and Deviant Behavior," in Rubenstein, E.A., Comstock, G.A., and Murray, J.P., eds., Television and Social Behavior. 4 vols, Rockville, Md.: National Institute of Mental Health, 1972. Vol. III, pp. 383-435.

NHK Radio and Television Culture Research Institute, "Studies of Children and Television in Shizuoka," Bulletin of the Radio and Television Culture Research Institute, various dates, 1965-1970.

Remmers, H.H., Horton, R.E., and Mainer, R.E., Attitudes of High School Students toward Certain Aspects of Television. West Lafayette, Indiana: Purdue University, 1953.

Robinson, J.P., "Television's Impact on Everyday Life: Some Cross-national Evidence," in Rubenstein, E.A., Comstock, G.A., and Murray, J.P., eds., Television and Social Behavior. 4 vols., Rockville, Md.: National Institute of Mental Health, 1972. Vol. IV, 1972, pp. 410-431.

Robinson, J.P., and Converse, P., The Impact of Television on Mass Media Usage. Paper read at Sixth World Congress of Sociology, Evian, France, September 1966.

Rubenstein, E.A., Comstock, G.A., and Murray, J.P., eds., Television and Social Behavior. 4 vols. Rockville, Md.: National Institute of Mental Health, 1972.

Schramm, W., Lyle, J., and Parker, E.C., Television in the Lives of Our Children. Stanford, California: Stanford University Press, 1961.

Stein, A., and Frederick, L., "Television Content and Young Children's Behavior," in Rubenstein, E.A., Comstock, G.A., and Murray, J.P., eds., Television and Social Behavior. 4 vols. Rockville, Md.: National Institute of Mental Health, 1972. Vol. II, pp. 202-317.

Szalai, et al., "The Multinational Comparative Time Budget Project," American Behavioral Scientist, December 1966.

Selected Percentiles for Scholastic Aptitude Test Scores (1966-67 through 1975-76)

June Stern
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

February 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

Table 1. Selected Percentiles for Scholastic Aptitude Test Scores, 1966-67 through 1975-76

Percentile	Testing year[a]									
	1966-67	1967-68	1968-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76
Verbal										
90th	618	617	613	613	604	598	590	589	583	579
75th	542	540	539	537	530	525	517	513	511	503
50th	459	460	456	455	450	443	438	433	432	424
25th	388	384	383	377	373	369	363	360	359	352
10th	327	324	321	316	312	309	305	302	298	289
Mathematical										
90th	645	647	642	641	640	635	634	637	630	630
75th	573	574	570	568	568	564	561	561	552	551
50th	490	491	488	487	483	481	480	474	469	468
25th	414	411	408	406	403	399	398	390	384	381
10th	352	345	344	336	336	331	330	324	322	314

a. Testing year is defined as extending from September through August.
Scores are aggregated without regard to educational level or repetition pattern within the defined year.

Table of SAT and Achievement Test Scores for Samples of Candidates Taking Achievement Tests 1966-67 to 1975-76

June Stern
Educational Testing Service

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

January 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

Table 1. SAT and Achievement Test Scores for Samples of Candidates Taking Achievement Tests, 1966-67 to 1975-76[a]

Year	Achievement			SAT-V		SAT-M		Achievement			SAT-V		SAT-M	
	N	M	SD	M	SD	M	SD	N	M	SD	M	SD	M	SD
	English Composition							American History and Social Studies						
1966-1967	4565	521	106	518	106	542	107	4702	503	113	511	108	517	107
1967-1968	4571	516	105	517	107	545	106	4533	500	106	512	109	522	109
1968-1969	4701	514	106	514	104	543	107	4634	496	104	513	104	525	106
1969-1970	3552	511	105	513	104	538	108	2901	482	106	506	105	516	104
1970-1971	5082	508	105	505	107	537	106	5098	488	102	504	109	519	104
1971-1972	6720	512	107	505	108	538	110	7101	487	103	501	105	519	109
1973-1974	4239	520	106	511	106	552	110	4049	497	105	503	106	527	108
1975-1976	5144	539	109	507	106	556	112	5819	495	107	505	107	536	111
	European History and World Cultures							French						
1966-1967	7946	541	108	541	109	516	104	4740	506	100	552	99	552	99
1967-1968	8149	535	107	542	111	520	105	4834	523	101	552	100	555	100
1968-1969	7235	540	108	541	110	526	103	3555	510	97	549	96	555	98
1969-1970	5826	530	108	539	111	522	104	4068	503	95	541	97	547	95
1970-1971	4841	527	107	542	113	523	104	5281	524	96	547	101	556	96
1971-1972	4746	532	108	549	112	528	108	8294	537	99	544	100	558	100
1973-1974	4040	521	108	544	114	530	108	4763	551	104	540	101	563	101
1975-1976	2576	531	111	561	113	555	112	5393	550	103	545	100	566	103
	German							Hebrew						
1966-1967	4925	500	100	561	97	581	99	851	589	99	554	110	575	111
1967-1968	5553	507	99	559	98	586	100	957	584	96	557	110	574	111
1968-1969	3779	508	95	555	95	585	99	989	569	97	552	110	575	110
1969-1970	4565	494	95	553	96	578	96	1025	573	103	540	112	560	106
1970-1971	4999	512	92	558	97	583	98	878	571	106	544	107	569	100
1971-1972	5691	531	101	556	98	584	99	932	563	101	557	104	570	110

(continued)

Table 1 (continued)

	Achievement			SAT-V		SAT-M		Achievement			SAT-V		SAT-M	
Year	N	M	SD	M	SD	M	SD	N	M	SD	M	SD	M	SD
	German							Hebrew						
1973-1974	5134	533	105	552	98	588	102	679	561	97	531	108	563	109
1975-1976	4075	538	104	552	101	592	103	493	565	99	536	103	558	110
	Latin							Russian						
1966-1967	6556	536	98	552	101	558	99	830	511	114	598	102	612	104
1967-1968	5858	534	101	555	104	566	102	917	511	110	601	98	617	102
1968-1969	5163	531	100	550	100	565	98	883	508	112	594	97	610	99
1969-1970	6842	523	105	550	100	562	96	891	524	132	574	112	590	106
1970-1971	3752	534	99	551	97	569	96	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1971-1972	3639	535	101	560	100	576	96	515	513	112	592	94	612	101
1973-1974	1795	537	103	556	101	582	100	224	555	137	582	106	614	107
1975-1976	917	519	98	573	100	591	100	277	514	115	571	109	596	113
	Spanish							Biology						
1966-1967	5140	482	102	522	101	532	99	4718	510	107	508	109	516	105
1967-1968	4813	490	102	523	100	538	101	4653	503	110	509	111	516	106
1968-1969	3610	494	100	518	96	534	97	4773	509	108	511	106	520	102
1969-1970	3785	479	97	513	96	529	96	3889	498	110	505	109	513	106
1970-1971	5183	507	100	518	100	536	98	5074	517	108	503	108	517	104
1971-1972	6503	511	98	513	101	534	102	6580	527	108	504	107	522	106
1973-1974	4690	538	110	509	104	537	107	3864	536	106	507	105	532	108
1975-1976	5511	529	106	508	104	540	107	4881	536	109	504	106	535	107
	Chemistry							Physics						
1966-1967	4679	552	99	530	102	593	99	5045	581	98	535	106	623	95
1967-1968	4578	556	106	537	106	599	100	5103	577	100	542	110	629	96

(continued)

Table 1 (continued)

Year	Achievement N	Achievement M	Achievement SD	SAT-V M	SAT-V SD	SAT-M M	SAT-M SD	Achievement N	Achievement M	Achievement SD	SAT-V M	SAT-V SD	SAT-M M	SAT-M SD
	Chemistry							Physics						
1968-1969	4877	564	104	531	101	601	96	5057	581	105	534	107	631	95
1969-1970	3774	559	105	528	103	597	98	3831	573	101	530	109	625	94
1970-1971	4555	563	106	523	104	592	100	4097	584	106	529	111	624	97
1971-1972	7618	563	106	524	107	593	105	6358	593	104	536	112	631	96
1973-1974	4592	565	110	526	107	602	103	4667	603	102	543	114	643	94
1975-1976	4520	563	103	527	108	611	108	4143	596	100	541	108	651	94
	Mathematics Level I							Mathematics Level II						
1966-1967	4605	549	103	507	102	554	102	4796	688	93	589	106	673	89
1967-1968	4536	543	100	510	107	556	103	5349	679	94	588	111	669	94
1968-1969	4637	538	101	507	103	554	101	4717	673	95	578	108	666	91
1969-1970	3714	538	104	506	103	548	100	3736	662	101	569	108	653	93
1970-1971	4839	537	94	498	102	543	99	5946	659	99	561	112	648	97
1971-1972	6788	540	100	500	105	543	102	7920	664	99	567	106	650	92
1973-1974	4172	541	97	498	101	548	101	4368	668	96	563	108	656	92
1975-1976	5174	547	100	492	104	549	104	5246	671	96	558	108	660	91
	Literature													
1966-1967														
1967-1968														
1968-1969	7332	531	111	538	111	511	104							
1969-1970	4736	525	109	530	110	504	101							
1970-1971	4309	522	109	529	112	506	103							
1971-1972	7775	533	109	528	112	505	107							
1973-1974	3933	532	109	522	111	506	105							
1975-1976	5637	528	112	524	115	512	110							

a. Based on data obtained in rescaling Achievement Tests annually through 1971-72 and biennially after 1972. The sample sizes do not reflect the relative total volumes for the various tests.

Motivational Factors in the SAT Score Decline

David G. Winter
Associate Professor of Psychology
Wesleyan University

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

March 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

INTRODUCTION

There is now unmistakable evidence that scores on the Scholastic Aptitude Test (SAT) and other tests of academic performance have declined among American secondary school students since 1963. In this paper I suggest some motivational factors that may have contributed to this decline, or at least have been involved in it. I shall not argue that motivation is the complete explanation or the only explanation, but rather that motivational factors may have played some role in the downward drift of scores. My analysis and suggestions are not based on a priori speculation, but are drawn from the accumulated evidence of more than thirty years of systematic psychological research on several important human motives. This paper assumes, of course, that the SAT score decline is at least to some extent "real," that it cannot be explained (or fully explained) in terms of artifacts such as changes in test items, changes in the composition of groups tested, less complete congruence between subjects taught in school and subjects covered in the test, or other nonmotivational factors such as the putative influence of changes in family size on intelligence (cf. Zajonc and Markus, 1975). These issues have been discussed by Harnischfeger and Wiley (1975), among others. Also outside the scope of this paper (but treated by others) is the role of values -- about education, success, the good life, etc. -- in SAT score declines.

The major premise of this paper is that performance on tests such as the SAT, like performance on the academic tasks predicted by SAT scores, is some joint function of ability and motivation. Ability is what the tests are designed to measure, and is probably related to test performance in a simple, positive, monotonic way: other things being equal, the higher the ability -- the higher the test score. The nature and effects of motivation are, however, more complicated. The general term "motivation" stands for the total combined effect of different specific motives, some of which might be positively associated with test performance (e.g., motives to achieve or excel), and some of which might be negatively associated with test performance (e.g., anxiety or distracting motives). Certain motives, such as a motive to please parents, might help or hinder test performance depending on whether the relevant "other" (e.g., parent) valued this performance. Finally, most motives probably have a nonmonotonic relationship to test performance. The variable effects of anxiety are the most familiar example: at lower levels, it can energize a lethargic student, but at higher levels it can "freeze" or inhibit performance.[1] For these reasons, it should be clear that there will be no simple answer to the two questions of how "motivation" as a whole affects test performance, and how motivational factors may be involved in the decline of SAT scores. In such circumstances it is well to proceed one step at a time, and to examine the possible separate effects of each motivational factor.

In this paper, therefore, I shall examine the research evidence and theory about three important human social motives, each of which could be expected to play some role in test performance and academic functioning: the motives for achievement, for affiliation, and for power.[2] Clearly it is not intended to be an exhaustive or complete review of all psychological perspectives on the issue. In the research tradition that I shall discuss, these motives are usually measured by scoring samples of spontaneous or operant thoughts, associations, and imagination, rather than by asking people directly about their motive levels. Since the two ways of measuring motives do not yield the same results, I shall use the technical names n Achievement, n Affiliation, and n Power to designate motives measured in this operant way.

The Achievement Motive

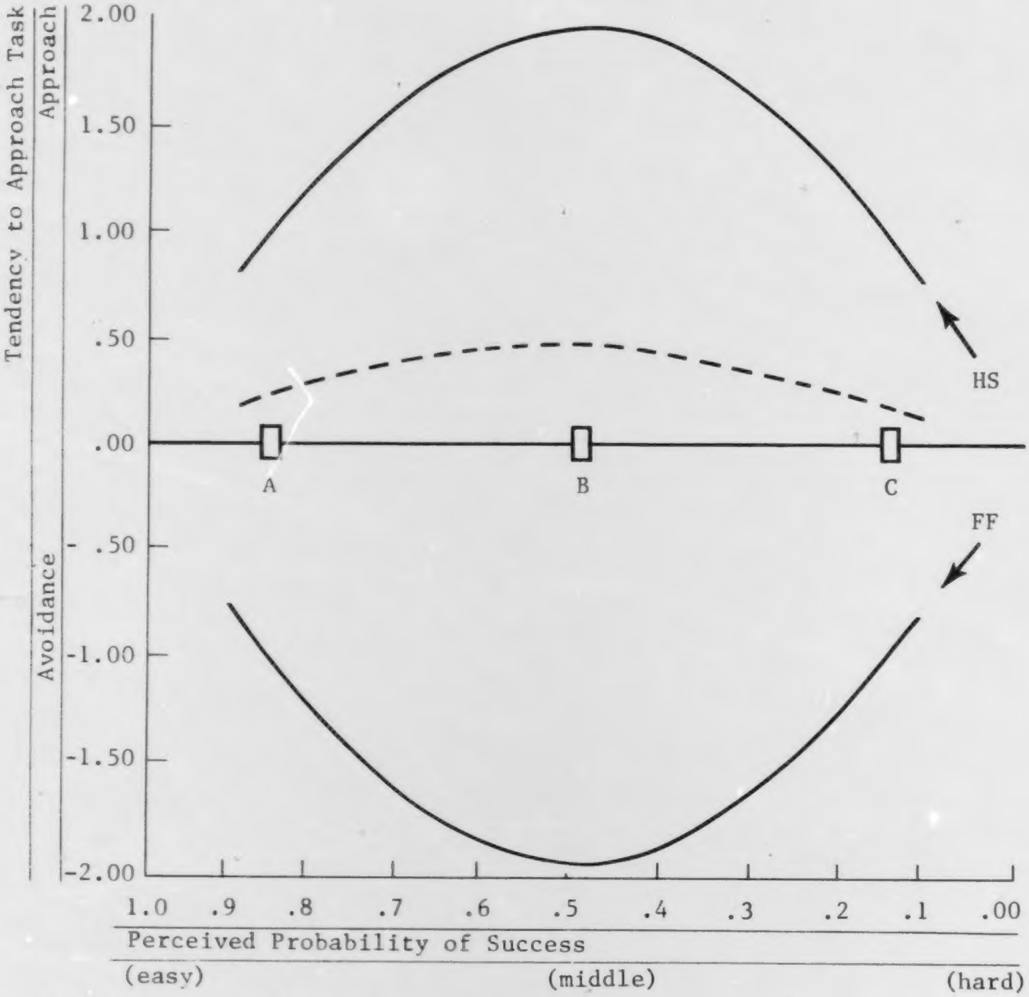
The major question of this paper can now be simply phrased: What do we know about the levels and characteristics of these three motives that might contribute to explaining the SAT-score decline? While it would seem obvious that the achievement motive (n Achievement), a concern with excellence and doing well, would be connected with academic achievement and therefore test performance, it consistently correlates much more highly with entrepreneurial or business achievement than with academic achievement. In order to explain this pattern, and to elucidate the ways in which n Achievement does influence academic and test performance, it is necessary to make a brief excursion into the theoretical model of the dynamics of action that has been worked out by Atkinson and his colleagues. The essential features of this model can be expressed in three propositions: (1) Achievement motivation involves two independent components -- an approach motive, or Hope of Success; and an avoidance motive, or Fear of Failure. The first component arouses action and leads us to choose and strive for goals that give us a feeling of excellence or having done well. The second component inhibits action and leads us to avoid choosing goals and acting in such a way that we are likely to fail. (2) Each component interacts in a multiplicative way with two further variables to produce action: the expectancy of the relevant goal (success or failure) and the value or incentive of that goal (amount of success feeling or amount of aversive failure feeling). For each component, this interaction can be represented by the following formula:

Tendency = Function (Motive x Expectancy x Incentive)

The resultant tendency to strive for a particular goal, or to choose a particular task, would then be the sum of the two component tendencies for that task -- the one positive, (approach tendency) and the other negative, (avoidance tendency). (3) At least in the domain of achievement goals, expectancy and incentive are inversely related -- i.e., as the perceived probability of success increases from .00 to 1.00, the subjective value or "achievement satisfaction" declines from 1.00 to .00. In like manner, as the probability of failure increases from .00 to 1.00, the aversiveness of failure declines. (It is, after all, not so bad to fail at an impossible task!) These three propositions are illustrated in Figure 1, which summarizes a good deal of research on level of aspiration, choice, persistence, and actual performance as a function of perceived task difficulty.

The top curve in Figure 1 (curve HS) shows the choice and approach behavior of a person high in (net) Hope of Success (arbitrarily set at 8), while the middle dashed curve shows the behavior of a person lower

Figure 1. Tendency to Approach a Task as a Function of Perceived Difficulty and Achievement-Related Motives



in net Hope of Success (arbitrarily set at 1). For each person, task B of middle difficulty is more attractive than task A (easy) or task C (hard); but for the high HS person the differential is quite substantial, while for the low HS person it is much smaller and could readily be overridden by other, extrinsic motive considerations (conformity, pleasing others, etc.). Notice, however, that the situation is exactly reversed for the high Fear of Failure person, as represented by the bottom curve (FF). Here moderate-difficulty tasks are the most aversive ones and therefore the ones most likely to be avoided. In fact, tasks of any difficulty are below the line of theoretical indifference, which suggests that FF people find any achievement task aversive, and would only choose and strive in the achievement domain if they are forced to or if other, extrinsic motives are in operation. Once in the achievement domain, they will choose (and work hardest at) tasks that are either very easy or very hard -- tasks A or C in Figure 1.

Finally, the expectancy of success, or perceived difficulty of a task, can change -- either through direct alteration (as when a supervisor or teacher adjusts a task so that it is harder or easier) or indirectly, through our experiences of success or failure. Tasks on which we continually succeed become perceptibly easier (a leftward move in Figure 1), while tasks at which we keep failing increase in perceived difficulty (a rightward move). Whether these changes will make the task more or less attractive -- whether the curve of tendency to approach is rising or falling -- does depend on the initial task difficulty; but it also depends on which curve it is -- i.e., whether the person's predominant motive component is Hope of Success or Fear of Failure. As an example, taking a very difficult task and making it easier (to a level of moderate difficulty -- e.g., C to B in Figure 1) will increase the tendency of HS-motivated people to choose it and do well at it; but it will make the task substantially less attractive to FF-motivated people. They will then be less likely to choose it than before or, having chosen it, will show a decrement in performance. To take another example, continued failure at a moderate task would make it less attractive to HS-motivated people, who would respond by lowering their aspirations to more "reasonable" levels; but the same continued failure would make the task more attractive to FF-motivated people, who would persist and even shift their aspirations upward in a way that appears counter to reason.

This model explains why HS-motivated people are likely to achieve and succeed in business: they select and strive hardest at tasks of moderate (versus extreme) difficulty, and they shift their aspirations "rationally" after success and failure. The model also suggests several ways in which the two components of achievement motivation should affect test performance and academic striving. (1) First, when students believe that a particular academic or test performance is instrumental to a long-term achievement goal (i.e., the particular task has associated with it a high perceived probability of success that it will advance the person along the path to a valued achievement goal) then that task becomes an "achievement task." It engages the two components of the achievement motive, as has been experimentally demonstrated several times by Atkinson and Raynor (1974). (2) Assuming that the two achievement-related motives have been engaged by the task, then aspirations, persistence and performance will vary with changing levels of perceived difficulty in the ways specified by the model.

This discussion suggests several different explanations for the decline in SAT scores over the last fifteen years. For the most part these explanations must stand as guides for further inquiry and the organization of what we already know, since direct evidence for any of them is lacking at this time.

1. Secular decline in n Achievement (Hope of Success), assuming that academic test performance is perceived as instrumental to attaining further valued achievement goals. While many observers have detected a decline in the achievement ethic, there has been little systematic testing of comparable groups of students, over time. Winter, McClelland, and Stewart (1977) did find that n Achievement levels declined sharply among freshmen men and women at an elite "Ivy" university between 1960 and 1974. A careful search of data archives of investigators who have tested high school and college students over the past 25 years would provide

further information on whether this was an isolated result or part of a secular decline in individual n Achievement. McClelland (1975, p. 344) found that n Achievement levels in American cultural documents increased continuously from the 1920s through the 1960s, which is the approximate beginning point of the SAT decline. More extensive year-by-year samples of such cultural documents (works of fiction, children's readers, hymns and popular songs, etc.) could be coded in order to pinpoint recent cultural trends in motive levels.

Apart from such overall national trends, it is also possible that n Achievement has declined among some socio-economic and ethnic groups (e.g., those whose children go to SAT-required colleges) while increasing among other groups (e.g., those who tend not to go to college, or who tend to go to colleges where the SAT is not required). Thus Veroff, Feld, and Gurin (1962) found that middle-aged, middle- and lower-income Roman Catholics with large families had extremely high levels of n Achievement when tested in a national sample survey in the mid-1950s. It may be that achievement motivation and perhaps academic performance has, as it were, "migrated" to this group and away from the traditional upper-middle class Anglo Saxon Protestant groups. This should be considered together with Horner's argument that college continues to be for some groups a means to social mobility but not so for other groups for which it was so in the past.

2. Secular increase in Fear of Failure. While Hope of Success and Fear of Failure are two independent components of the achievement motive, any increase in Fear of Failure should, other things being equal, have effects on academic choice and performance that are similar to the effects of declining Hope of Success. In Figure 1, either such change would be a shift from the HS line to the FF line. Tasks that used to be readily approached and persistently worked on (middle-range tasks) would now be most strongly avoided. In addition, the whole domain of achievement behavior would become to some extent aversive. In the research tradition that is the focus of this paper, Fear of Failure is often measured by questions such as the following, which reflect debilitating anxiety: "Nervousness while taking an exam or test hinders me from doing well." and "The more important the examination, the less well I seem to do." Systematic evidence has yet to be assembled as to whether students agree with items of this sort more today than they did twenty years ago. Nevertheless, it is reasonable to suppose that they do because of increasing pressure for every kind of accomplishment and success in our culture.

Personal, social, and economic stakes are higher in the academic arena today. Increasingly, one's ultimate position (job, status) is based on one's performance, rather than on one's ascribed characteristics such as race, sex, or social class. In a meritocracy of this sort, performance standards are continually increasing. Increasing levels of anxiety are, on this hypothesis, a natural result of the higher standards and heightened competition (especially in a constrained economy).

3. Decline in the belief that academic and test performance is instrumental in attaining further valued achievement goals. Such a change in students' beliefs would have the effect of disengaging their achievement motives from academic performance, and reengaging them in whatever other actions were perceived as relevant to long term achievement goals. With this motivational support withdrawn, performance would tend to decline.

Archives of public opinion polls among high school and college students could be examined in order to test this explanation. The further corollary, that the correlation between n Achievement and high school academic performance should have decreased significantly over the years of the SAT score decline, could perhaps be tested by drawing together findings from particular studies that assessed both motive and performance in different, specific, years during this time period. Meanwhile, anecdotal and observational support for this explanation can be drawn from the assertions of Bird (1975) and numerous others who have questioned the relevance of academic performance to any number of later vocational and other achievement goals. The current debate about the relative merits of a liberal arts education and training of "salable skills" is pertinent here.

4. Changes in secondary schools that, interacting with HS and FF, decrease the approach tendency toward academic tasks. This complex of explanations follows from the model in Figure 1. Several possibilities suggest themselves. For example, we know from past research that shifts toward homogeneous ability grouping in secondary schools should increase academic effort among HS motivated students but decrease it among FF motivated students (see O'Connor, Atkinson, and Horner, 1966). If academic tests and tasks used to be perceived as very difficult, then any secular trend toward making school easier would increase performance among HS students but decrease it among FF students. On the other hand, if school were perceived as only moderately difficult before the SAT score decline, then making things easier should have the opposite effects. (The same explanation applies, *mutatis mutandis*, to increasing school difficulty.) One might be tempted to discard this explanation, since the presumed opposite effects for HS and FF students ought to cancel each other out. However, secular changes in both school difficulty and relative predominance of HS or FF could produce an overall decline in performance. Birney, Burdick, and Teevan (1969) found that themes of "Hostile Press" (a variable conceptually and empirically related to Fear of Failure) predicted grades among junior high school, high school, and college students. If these results are generally true, then one is led to a most surprising conclusion. Educational changes that made tests and academic tasks "moderately difficult" instead of "very difficult" -- changes that seem reasonable, progressive and humane -- would have the paradoxical effect of making tests and school less attractive, unless further steps were taken to engage HS.[3]

5. Increase in Fear of Success. The two components of achievement motivation that have so far been discussed -- Hope of Success and Fear of Failure -- are "rational" in that they are congruent with traditional cultural values about the achievement domain. Recently, however, Horner (1974a, 1974b) has identified another component labelled *Fear of Success*, which seems to run counter to our cultural prescriptions. Horner argues that for some people, achievement and success are punished -- most notably through withdrawal of love, affiliation, and peer approval. Success for such persons may thereby acquire a negative valence or incentive value -- it becomes something to be avoided. Once developed, Fear of Success tends to lower performance levels in competitive achievement situations. Women and minority group members are the traditional and most striking examples of Fear of Success inhibiting performance. However, in recent years there is at least some evidence that Fear of Success has

increased among white middle-class males also.[4] Perhaps this is a psychological manifestation of the more general "counter-culture" revolt against what is viewed as the "excessive" or "inhumane" achievement and success emphasis of traditional American culture. For whatever reasons success might have become something to fear, increased Fear of Success (as measured in more formal and operational terms) would lower performance precisely in the kind of competitive testing situation of the SAT as well as in a variety of other academic situations.

The Affiliation Motive

People who score high in *n* Affiliation seek warm, friendly, and cooperative relationships with others, although they can be suspicious and defensive if they think others are rejecting them. *N* Affiliation can sometimes act as a powerful stimulus to academic striving, as in the case of the student who does well in order to please the teacher or parents. There is some research evidence that in a classroom environment rich in affiliative cues and incentives, *n* Affiliation predicts performance (McKeachie, 1961). On the other hand, in such cases *n* Affiliation is always being deflected from its primary goal of friendly interaction. Over time, it would be difficult to keep the motive on the "academic" track with alternative behaviors that are more efficient in attaining the goal readily available. Thus while the affiliation motive may produce good social relationships and a cooperative, friendly attitude in the classroom, it is a motive that is extrinsic to academic performance as such. Under some circumstances it could be engaged to support academic performance, but this support is not likely to be sustained if alternative sources of affiliation are readily available.

6. *Increases in N Affiliation*. In the late 1960s love and intimacy emerged as strong themes in American popular culture. It is surely plausible to argue that the increased prominence of these themes is one manifestation of a general increase in *n* Affiliation, while decreased emphasis on school achievement and lower test scores is simply another sign of the same trend. On the other hand, there is little direct evidence of higher *n* Affiliation levels. Winter, McClelland, and Stewart (1977) found no shifts, and McClelland (1975, p. 338) actually found that *n* Affiliation coded from cultural documents decreased from the 1950s to the 1960s.

7. *Educational policy changes that have increased the role of n Affiliation*. Even though average levels changed, the affiliation motive could have an indirect effect on academic and test performance in conjunction with changes in educational policy and practice. Many of the educational reforms of the past fifteen years, for example, emphasized the role of affiliation in the academic process. General concern for students' social adjustment increased, and specific steps were taken to decrease the distance between teacher and student, and to reduce hierarchy in the classroom. In motivational terms, the effect of these changes would be to engage *n* Affiliation, to have this motive carry more of the load of supporting academic behavior. Given the premise that schools need all the motivational help they can get, such a broadened motive base for academic effort seems to be a wise step. On the other hand, trying to engage *n* Affiliation in the classroom does involve some risk. The world of the

adolescent is full of other incentives and "outlets" for affiliation; most of them are available with less effort and greater certainty than is school. Put another way: in emphasizing affiliative rewards, a school may be trying to broaden its appeal, but it is then competing in a league where it is at a relative disadvantage (unless it renounces achievement standards completely and becomes just another "peer group institution"). Thus, the increased school emphasis upon affiliation incentives might, then, have the paradoxical effect of distracting students -- of driving them to seek these incentives outside of the school. In any case, to the extent that a school reduces achievement incentives in favor of other goals (in either relative or absolute terms), it is sacrificing one of its most potent and specific incentive systems.

The Power Motive

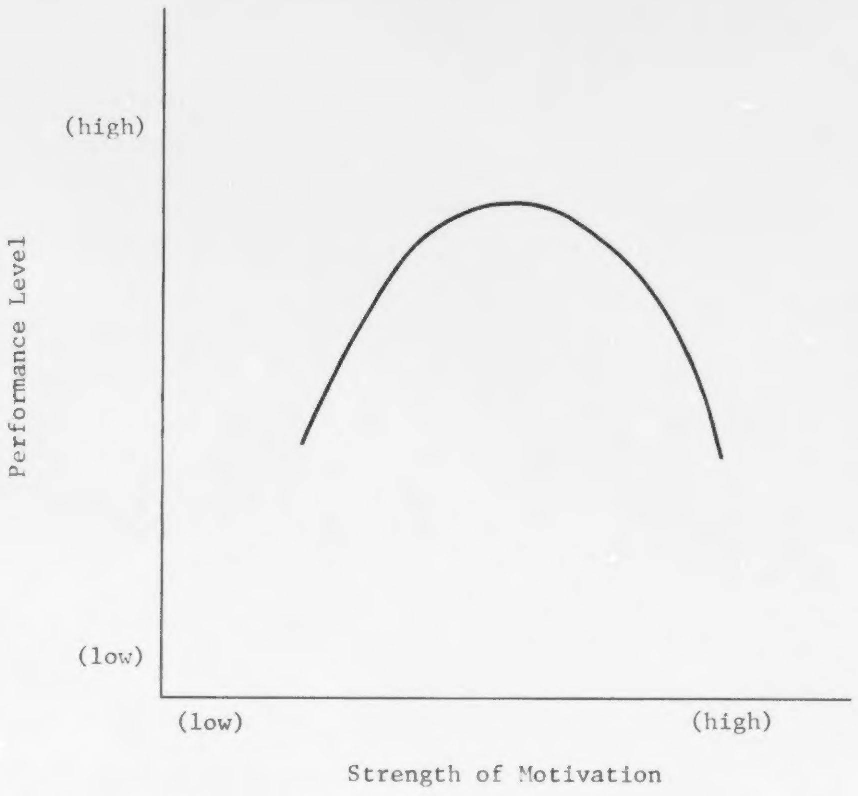
People who score high in n Power are concerned with having impact on others and gaining prestige in the eyes of the world. They seek and get formal social power, but they affect other people in a variety of less formal and less socialized ways, such as by taking extreme risks and engaging in profligate drinking and sexual behavior. While under certain circumstances, when the classroom environment is rich in power cues, the power motive may facilitate performance (see McKeachie, 1961), in general n Power does not seem to be a motive that readily supports academic performance. Learning and test performance are in some sense individual activities and the adolescent world is simply too full of readily available, more attractive, power incentives.

8. Increases in N Power. An obvious motivational hypothesis to explain the decline in SAT scores is, therefore, an increase in levels of the power motive among students. Certainly "power" was a central concern of the 1960s and 1970s: the emphasis upon "vigor" in the Kennedy years; the struggle for and against political power on issues such as Vietnam, civil rights, and other allied "movements"; the prominence of powerful multinational corporations; the increased emphasis upon violence in the media, sports, and fiction (For example, the James Bond series); and finally, the renewed interest in the occult as a source of secret and esoteric power. In line with these aspects of American cultural life, McClelland found a sharp increase in n Power levels in cultural documents from the 1950s to the 1960s. Winter, McClelland, and Stewart (1977) found that while total power motive scores did not increase significantly from 1960 to 1974, Fear of Power (the avoidance component of the motive) did increase. The overall picture is far from clear, and more evidence needs to be gathered on this point. Meanwhile, it is plausible to argue that increased concerns with power, impact, and prestige at all levels of society have drawn concern, energy and effort away from academic and test performance. Power is, then, another distraction from the academic situation.

Motivational Overload?

A final motivational hypothesis derives not from the study of any particular motive, but rather from the total sum of all motivational influences.[5] One of the oldest and best well established principles of psychology is expressed in the Yerkes-Dodson law: that as levels of motivation (or arousal) increase, efficiency and effectiveness of performances first increase but then drop. Performance is optimal at some middle level of motivation (the exact point, of course, depending on the nature of the task and on individual differences). This relationship is shown in Figure 2.

Figure 2. Relation of Motivation Level to Performance: The Yerkes Dodson Law



How might such a principle explain the decline in scores? First, let us assume that academic performance is ordinarily based on the achievement motive. The increased emphasis on the instrumental importance of education (and test scores) over the past two decades has, arguably, increased achievement motive arousal in academic settings. At the same time, moreover, educational changes have engaged other motives (e.g., affiliation and power) to support academic performance. As a result, the total sum of motivation directed toward academic and test performance is greatly increased. (In Figure 2, this is represented by a move to the right along the curve.) The effect on performance of such an increase in total motivation will vary from student to student, depending on individual differences in motive levels and other variables. As the total level of motivation increases, however, performance will begin to decline in more and more students, "over motivation" being the cause.

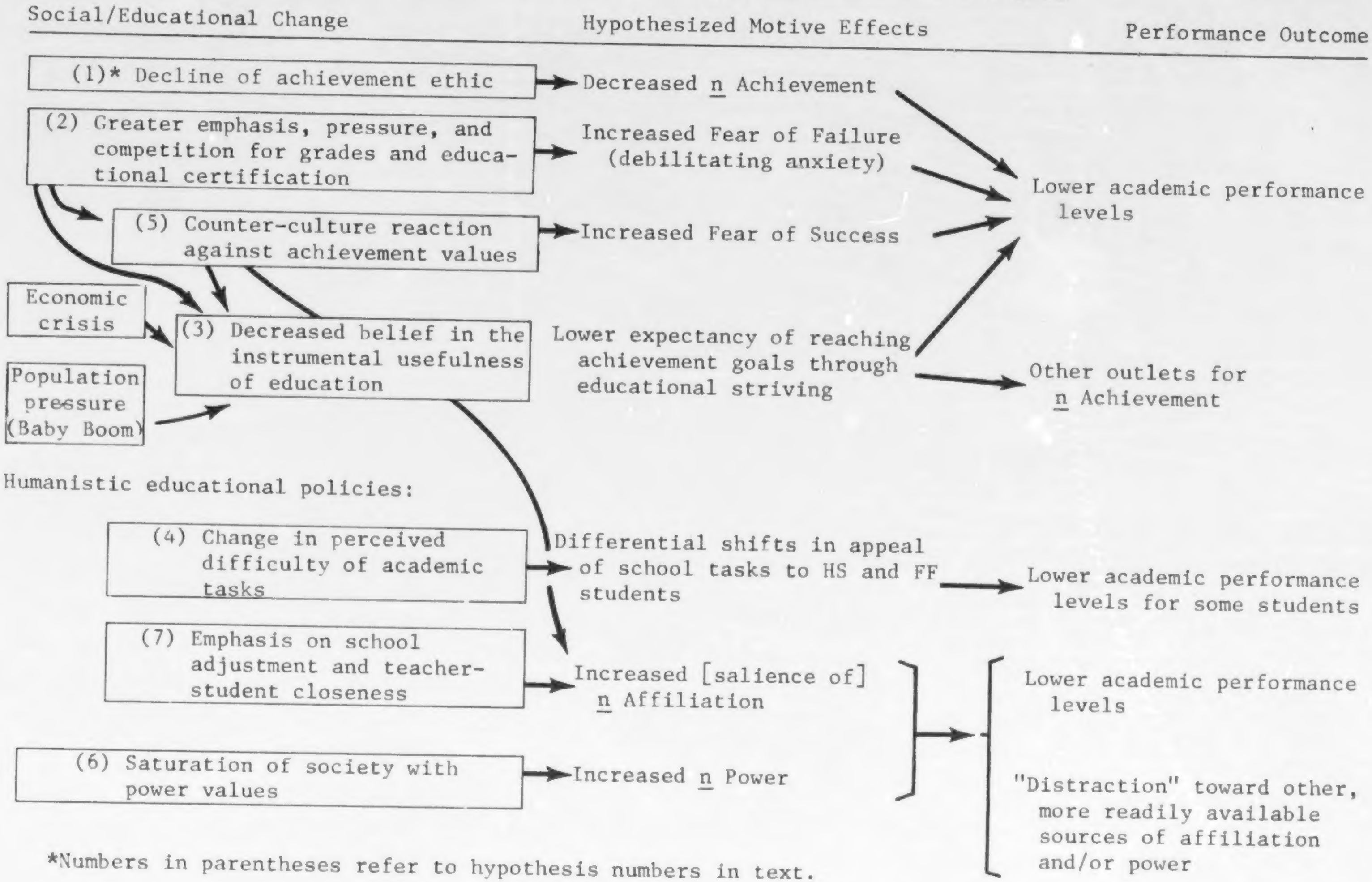
This hypothesis suggests that what has declined is performance on a standardized test, and not ability as such or even academic achievement overall, since cumulative academic achievement depends upon amount of time spent in academic work as well as on efficiency level while at work. Time spent in an activity should be a direct and not curvilinear function of motivation levels for that activity.[6] While such a prediction from the hypothesis is clear, it is also difficult to evaluate because of our uncertainty about trends in cumulative academic achievement. To be sure, average grades have increased, but many educators would argue that this is because of "grade inflation" rather than improved performance. Yet the best measures of performance -- the most comparable across time and the least subject to inflation -- are the standardized tests, for which students are hypothetically overmotivated.

The Uses of a Motivational Analysis

This paper has suggested certain ways in which particular human social motives may have played a role in the recent decline of SAT scores. Several cultural trends have emerged in American society over the past fifteen years: emphasis on academic achievement in the "post-Sputnik" era, the revolt of the counterculture, the entry of the "Baby Boom" children into the work force, and the crisis of values in Vietnam and Watergate. Related to these broader trends, certain specific changes appeared in the educational system: school as a focus or crucible for broader social, cultural, and political views and conflicts; trends toward "open," "alternative," and "humanistic" education; questioning of traditional subjects and methods of instruction. These broader social forces, in turn, have direct and indirect effects on the levels and functioning of particular human social motives such as the various components of achievement motivation, and the affiliation and power motives. Some of the hypotheses outlined above suggest that actual motive levels went up or down with corresponding effects on academic performance. Other hypotheses focus on the ways in which broad social forces may have changed beliefs or expectancies about the nature and instrumental value of academic work; these changes affect academic performance by interacting with more stable motive levels. Social trends and educational changes "cause" changes in behavior not by themselves, but rather in combination and interaction with individual persons. Motives are, in this view, one important medium through which social events and trends are transmitted down to the level of changed behavior of individuals and aggregates of individuals. Figure 3 charts the analysis and shows how certain hypothesized motive changes and motive-related effects can be derived from the broader social forces. In some cases, the social forces themselves interact, as shown in the figure.

Some of the suggested hypotheses overlap, while others are mutually contradictory. Further evidence will be needed in order to decide among the hypotheses. One important source of relevant data is the numerous studies carried out by a variety of investigators over the past 20 years in which adolescents were given SATs or similar tests. We have used data

Figure 3. Relations of Social Change, Individual Motives, and Academic Performance


Social/Educational Change
Hypothesized Motive Effects
Performance Outcome
(1)* Decline of achievement ethic
Decreased n Achievement
(2) Greater emphasis, pressure, and competition for grades and educational certification
Increased Fear of Failure (debilitating anxiety)
Lower academic performance levels
(5) Counter-culture reaction against achievement values
Increased Fear of Success
Economic crisis
Population pressure (Baby Boom)
(3) Decreased belief in the instrumental usefulness of education
Lower expectancy of reaching achievement goals through educational striving
Other outlets for n Achievement
Humanistic educational policies:
(4) Change in perceived difficulty of academic tasks
Differential shifts in appeal of school tasks to HS and FF students
Lower academic performance levels for some students
(7) Emphasis on school adjustment and teacher-student closeness
Increased [salience of] n Affiliation
(6) Saturation of society with power values
Increased n Power
Lower academic performance levels
"Distraction" toward other, more readily available sources of affiliation and/or power

*Numbers in parentheses refer to hypothesis numbers in text.

from one such study in this way (Winter, McClelland, and Stewart, 1977), noting changes in motive levels of college freshmen over fourteen years. This study is quite fragmentary -- for only one picture was used in both tests, and the subjects were highly selected. By systematically noting the year of testing, age of sample, and average motive level, it would be possible to document and track secular changes in motive levels that would support particular hypotheses and rule out others. (It would be necessary, of course, to take into account and track separately each separate picture or verbal stimulus in order to insure comparability.) Moreover, where investigators had gathered information about school grades and academic test performance, this information could also be noted in order to plot systematically any changes in the pattern of relationships among motive score, ability, and performance.

Another important source of relevant data will be the forthcoming research designed to replicate the 1950s study in which the Thematic Apperception Test was administered to a national sample survey (Veroff, Atkinson, Feld, and Gurin, 1960; see also the Veroff, Feld, and Gurin, 1962, study cited above). When completed, this research will pinpoint which motives changed in which population subgroups during the 20 year interval between the 1950s and the 1970s (Joseph Veroff, personal communication).

Finally, once data of this sort have been collected, it would also be possible to analyze the broad question of motivational elements in the SAT score decline into four more differentiated and specific problems: (1) Changes in the motives of students relevant to taking the SAT test itself; (2) Changes in the motives of students relevant to activities preparatory and instrumental to the SAT-test (taking hard courses, studying hard, etc.); (3) Changes in the motives of teachers and school administrators relevant to the way in which they do their jobs; and (4) Changes in motives in the public constituencies that elect school boards, vote taxes, and to whom the whole educational system is, in the end, accountable.

The major purpose of this paper has been to suggest ways in which theory and research on human motivation and specific human motives can broaden the explanation of the SAT score decline phenomenon, and to suggest how further data to support these explanations can be gathered from existing research files scattered among several investigators across the country. Once established and documented, a motivational explanation could also be of use in order to arrest and reverse the decline. There now exists a considerable body of tested theory, practice, and educational "technology" about changing motives (especially in adolescent school populations) and changing classroom climate in order to enhance motivated performance (McClelland and Winter, 1969; Kolb, 1965; Alschuler, Tabor, and McIntyre, 1970; McClelland and Alschuler, 1971; deCharms, 1976). Drawing upon these projects, it should be possible to design an educational environment in which students' motives function more effectively -- to their fullest potential in both test performance and long term academic achievement.

Notes

1. This phenomenon, known technically as the Yerkes-Dodson law, will be discussed below under the heading of "motivational overload."
2. For further information about the nature and measurement of these motives, see: Atkinson (1958); Atkinson and Feather (1966); Atkinson and Raynor (1974); McClelland (1961); McClelland (1975); McClelland and Steele (1972); Winter (1973); Winter and Stewart (in press).
3. See Alschuler, Tabor and McIntyre (1970) and McClelland and Alschuler (1971) for a discussion of educational changes that engage HS.
4. Personal communication from Professor Lois W. Hoffman, University of Michigan.
5. Personal communication from Professor John W. Atkinson, University of Michigan.
6. J. R. Sawusch, in Atkinson and Raynor (1974 pp. 425-438).

Bibliography

Alschuler, A. S., Tabor, D., and McIntyre, J., *Teaching Achievement Motivation*. Middletown, [illegible]t.: Educational Ventures, Inc., 1970.

Atkinson, J. W. (ed.), *Motives in Fantasy, Action and Society*. Princeton, N.J.: D. Van Nostrand Company, 1958.

Atkinson, J. W., and Feather, N. T., *A Theory of Achievement Motivation*. New York: John Wiley & Sons, Inc., 1966.

Atkinson, J. W., and Raynor, J. O., *Motivation and Achievement*. New York: V. H. Winston & Sons, Inc., 1974.

Bird, C., *The Case Against College*. New York: David McKay Company, Inc., 1975.

Birney, R. C., Burdick, H., and Teevan, R. C., *Fear of Failure*. New York: Van Nostrand Reinhold Company, 1969.

deCharms, R., *Enhancing Motivation: Change in the Classroom*. New York: Irvington Publishers, Inc., 1976.

Harnischfeger, A., and Wiley, D. E., *Achievement Test Score Decline: Do We Need to Worry*? Chicago: ML-Group for Policy Studies in Education, CEMREL, Inc., 1975.

Horner, M. S., "The Measurement and Behavioral Implications of Fear of Success in Women." In J. W. Atkinson and J. O. Raynor (eds.), *Motivation and Achievement*. New York: V. H. Winston & Sons, Inc., 1974 (a).

Horner, M. S., "Performance of Men in Noncompetitive and Interpersonal Competitive Achievement-Oriented Situations." In J. W. Atkinson and J. O. Raynor (eds.), *Motivation and Achievement*. New York: V. H. Winston & Sons, Inc., 1974 (b).

Kolb, D.A., "Achievement Motivation Training for Underachieving High-School Boys," *Journal of Personality & Social Psychology*, 1965, *2*, 783-792.

McClelland, D. C., *The Achieving Society*. Princeton, N.J.: D. Van Nostrand Company, 1961.

McClelland, D. C., *Power: The Inner Experience*. New York: Irvington Publishers, Inc., 1975.

McClelland, D. C., and Alschuler, A. S., *Achievement Motivation Development Project, Final Report*. Project #7-1231, Grant #0-8-071231-1746. Washington: United States Office of Education, Bureau of Research, 1971.

McClelland, D. C., and Steele, R. S. (eds.)., Human motivation: A Book of Readings. Morristown, N.J.: General Learning Press, 1972.

McClelland, D. C., and Winter, D. G., Motivating Economic Achievement. New York: Free Press, 1969.

McKeachie, R., "Motivation, Teaching Methods, and College Learning." In M. R. Jones (ed.) Nebraska Symposium on Motivation 1961. Lincoln, Neb.: University of Nebraska Press, 1961.

O'Connor, P. A., Atkinson, J. W., and Horner, M. S., "Motivational Implications of Ability Grouping in Schools." In J. W. Atkinson and N. T. Feather (eds.), A Theory of Achievement Motivation. New York: John Wiley & Sons, 1966.

Veroff, J., Atkinson, J. W., Feld, S. C. and Gurin, G., "The Use of Thematic Apperception to Assess Motivation in a Nationwide Interview Study." Psychological Monographs, 1960, 74, 12 (Whole number 499).

Veroff, J., Feld, S. C., and Gurin, G., "Achievement Motivation and Religious Background, American Sociological Review, 1962, 27, pp. 205-217.

Winter, D. G., The Power Motive. New York: Free Press, 1973.

Winter, D. G., McClelland, D. C., and Stewart, A. J., College and Competence: Evaluating the Liberal University. Manuscript in preparation, Wesleyan, Harvard, and Boston Universities, 1977.

Winter, D. G., and Stewart, A. J., "Power Motivation." In H. London and J. Exner (eds.), Dimensions of personality. New York: John Wiley & Sons, Inc., in press.

Zajonc, R. B., and Markus, G. B., "Birth Order and Intellectual Development." Psychological Review, 1975, 82, pp. 74-88.

List of Hypotheses Advanced to Explain the SAT Score Decline

Yvonne L. Wharton
College Board

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

February 1976

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $2 per copy, as well as a price list for all the appendixes, can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

In the course of the revelation of the score decline, a great number and variety of explanations were suggested. Some were triggered by the first publicity given to the decline nearly four years ago. Many more surfaced, however, as a direct result of newspaper stories that followed Mr. Marland's statement on the score decline in his report delivered at the 1975 Annual Business Meeting of the College Board, which included the announcement of the appointment of the Advisory Panel. While most have been in the form of letters to Mr. Wirtz or Mr. Marland, a number have also appeared in newspaper stories and magazine articles. Some have concentrated on one theme while others have ranged across a number of theories.

In order to give the Panel some feel for the many types of explanations that have been offered, a listing is attached in a format or framework designed to group them in a meaningful way. The task of organizing the material has been complicated by problems of interrelatedness and overlap among many of the explanations offered. Efforts to accommodate subtle differences in perspective and emphasis among individual observers have led in some instances to the multiple listing of an item. Wherever possible the phrasing of the explanation has been taken from the original submission or source.

CHANGES IN THE SCHOOLS

E-I. Curriculum

1. neglect of the three Rs (educational innovation at the expense of basic skills)
2. increased emphasis on aural-oral communication skills (use of audio-visual equipment)
3. decreased emphasis on reading and writing skills
4. effects of "new math"
5. decline in modern foreign language study (Source: Modern Language Association)
6. decreased enrollments in traditional, basic academic courses (Source: Wiley report)
7. reading programs that do not emphasize phonetics, vowels, consonants, etc.
8. inferior textbooks (publishers' concern with form rather than content. Sources: newspaper article; Hudson Institute report)
9. stress on independent study
10. open classrooms (where students choose activities in non-graded groups and work with little teacher interference)
11. education as entertainment concept (widening variety of nonchallenging, nonacademic courses that entertain more than they educate. Sources: newspaper articles; U.S. News & World Report article)
12. different curriculums followed by different ethnic groups (black students in one California district are enrolled in algebra and geometry courses at half the rate of white students in the same district. Source: letter from provost, University of California at San Diego)

E-II. Institutional Policies

1. automatic promotion of students (whether or not they have mastered course content)
2. reduced emphasis on grades
3. failure to separate students into fast/slow classes on the basis of their abilities
4. less homework
5. growth in school size (children benefit from smaller, more personal neighborhood schools. Source: letter from parent)
6. "edifice complex" with regard to school facilities (delusion that better schools require shinier physical plants, grander stadiums, fancier equipment, ever more sophisticated computers, laboratories and machines. Source: Rep. Jim Wright of Texas)
7. system of thousands of independent school boards that makes it impossible to set or demand national achievement standards. Source: letter from teacher)
8. increased retention in high school (more low-achieving students now continue schooling until graduation, with a resultant dilution of the educational experience for all. Source: Wiley report)

E-III. Teachers

1. fewer contact hours between pupils/teachers (encroachment of labor movement on teaching profession)
2. greater reliance on clerical aides
3. employment of inexperienced teachers in the K-6 grades (because of low school budgets. Source: letter from parent)
4. decline in the educational quality of teachers (poorly educated teachers whose own low level of academic achievement adversely affects their students' performance. Sources: newspaper article; letter from teacher; Hudson Institute report; letter from H. Rouse. See also Family, I.7)
5. egalitarian teachers and supervisors (who are willing or unable to enforce academic standards)
6. influence of the "soft pedagogical left" (which believes expressiveness is an adequate substitute for thinking and knowing and which views leniency as a kindness to the underprivileged. Source: Dean, Livingston College)
7. more married female teachers (who have more distractions, less dedication to teaching than older, single women. Source: letter from parent)
8. influx of male teachers and principals in the 1960s into traditionally female domain (with a concomitant relaxation in school discipline. Source: letter from former teacher)

E-IV. Students

1. decreased motivation of the test taker in an era of decreasing competition for college admission
2. negative attitudes toward education (rejection of traditional academic courses as not relevant)

3. decline in competitiveness
4. decreased test taking among abler students (talent and ability of brighter students being channeled into nontraditional, nonmeasurable activities)
5. truancy (increasing numbers of students are being absent from school, thereby decreasing the average amount of schooling per pupil and further burdening the teaching process. Sources: newspaper article; Wiley report)
6. undisciplined students (discipline problems have occupied teachers with keeping order in the classrooms and left less time for instruction. Sources: letters from parents; newspaper article)
7. alcohol and drug problems (Wiley report)
8. alarming increase in youth crime (Wiley report)

CHANGES IN SOCIETY

S-I. Family

1. permissiveness, abdication of parental role (parents' refusal to make selves final authority on good/bad, right/wrong)
2. parental coddling, indulgence (parents and teachers who seek to absolve children from difficult or unpleasant tasks)
3. women's liberation:
    a) more working mothers
    b) fewer women willing to be "professional mothers" (with the result that nutritional needs of children not being met at home; guidance, supervision of children increasingly shifted to schools and other outside agencies)
4. higher divorce rate (more single-parent families)
5. changes in expectations for education (upward and downward social mobility)
6. effects of family configurations on intellectual achievement (mean birth rank over years appears to correlate with SAT scores. Source: Zajonc)
7. "baby-boom" phenomenon (impact of the postwar baby boom on school enrollments; pupils who have been earning lower scores on the SAT may be those who received most of their education in schools affected by rapidly increasing enrollments. Source: letter from H. Rouse. See also Teachers, III.4)

S-II. Religion

1. growing rejection of traditional Western religions and concomitant search for meaning and relevance in other areas (religions of the East, drug-related religions, mysticism, witchcraft, astrology)
2. preference for fantasy over reality (celebration of the "ideology of irrationalism" in which knowledge is attained through "intuition, inspiration and revelation." Source: L. Trilling)

S-III. Civil Rights

1. increased encroachment of the courts, including busing and other efforts to increase minority integration (with resultant upheaval in secondary school programs and attendance)
2. increasing legislative involvement in education (legislatures have tied the hands of school personnel with respect to discipline. Source: newspaper article; letter from parent)
3. women's liberation (cf. Family)
4. greater access to higher education (the more accessible a college education, the less are high school students motivated to do well)

S-IV. Crisis of Values

1. impact of the counterculture (with its hostility to reason, to science, to technology, to industry, to the work ethic)
2. revolution in values, including a decline in the Protestant ethic (in an advanced, technically proficient democracy, the idea of success, in a certain sense, the idea of work itself, is growing obsolete)
3. subversive political activity (activity of Communists and organized crime in promoting pornography; the "grabbing of political power by the NAACP, NEA, CLU, and others. . .[so that] public school children are misguided through negative education. . .The intention is to groom them in total disgust with the nation, family, and self." Source: letter from parent)
4. rise of egalitarianism (See also National Priorities, below)

S-V. National Priorities

1. national commitment to widespread diffusion of learning (with its tendency to bring education down to lowest common denominator)
2. Viet Nam War (its end has reduced motivation for college attendance to avoid draft)
3. reduced international commitment (less emphasis on foreign language instruction)
4. use of schools to correct social ills of nation (related to Civil Rights, which see)
5. open admissions (related to Civil Rights, which see)

S-VI. Economic

1. taxpayer rebellion (resulting in tighter school budgets)
2. employment situation (which has eroded the certainty that a college education will necessarily lead to a job)
3. rising college costs (may have discouraged some superior students from applying to expensive schools that require higher SAT scores)
4. open admissions to college (in part as a result of over-expansion in higher education)

S-VII. Labor Movement in Education

1. collective bargaining agreements (which may result in teacher contracts protecting the incompetent and limiting the creativity of teachers)

S-VIII. Technological Changes

1. impact of television:
    - a) less time spent in reading, conversation
    - b) children's auditory skills honed at expense of visual
    - c) daytime school effort erased, overwhelmed by after school television viewing (with its silliness, violence)
    - d) television viewing makes it more difficult for children to relate to the printed word
2. increased use of other electronic communication
3. decline in the importance of the printed media (except flip charts)
4. information glut
5. health-related considerations:
    - a) increased use of chemical additives in foods
    - b) side effects of tranquilizers and other drugs
    - c) effects of anesthesia used in childbirth

CHANGES IN THE POPULATION

1. more educationally disadvantaged
2. fewer able students going directly to college
3. decline in repeaters
4. repeaters today less able
5. pool of talented students being shared with ACT
6. migration of more capable females to the technologies
7. post-World War II baby population unusual in a number of respects (family size, structure, SES)

PROBLEMS WITH THE TESTS

1. scaling and equating
2. lack of "linguistic clarity"
3. no longer relevant to what is being taught
4. cultural bias